



Universidade do Minho
Escola de Engenharia

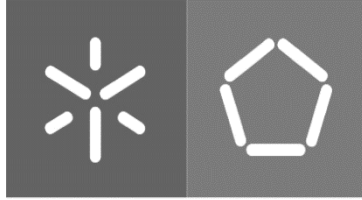
Hao Huang

Implementação de algoritmos de *TensorFlow*[™] para detetar Patologias cardíacas

Dissertação de Mestrado em Engenharia Eletrónica Industrial
e Computadores

Trabalho efetuado sob a orientação do
Professor Doutor Jorge Cabral

Janeiro de 2019



Universidade do Minho
Escola de Engenharia

Hao Huang

**Implementação de algoritmos de *TensorFlow™*
para detetar Patologias cardíacas**

Dissertação de Mestrado em Engenharia Eletrónica Industrial
e Computadores

Trabalho efetuado sob a orientação do
Professor Doutor Jorge Cabral

Janeiro de 2019

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

Agradecimentos

A realização desta dissertação de mestrado só se tornou possível pelos indispensáveis ajudas de várias pessoas. Deixo-lhes aqui o meu sincero agradecimento:

Em primeiro lugar, gostaria de agradecer aos meus pais por acreditarem sempre em mim e por todo incentivo que me deram não só durante este percurso académico, mas deste sempre.

Um obrigado ao orientador Jorge Cabral, pela disponibilidade que mostrou durante todo o decorrer do trabalho e pelas sugestões e conselhos no sentido de melhorar o trabalho.

Aos amigos e colegas por todo o apoio, pelas suas ideias, e pela sua constante disponibilidade. Sobretudo, aos meus amigos Nuno Silva, Bruno Amaro, Zé Oliveira, Filipa Araújo e Luis Miguel, um obrigado especial por todo o apoio, ajuda e por toda a motivação mesmo nos momentos de maior desânimo.

À minha namorada quero agradecer por todo o apoio e incentivo. Obrigado por me aturares nos momentos mais difíceis da minha vida.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração. Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Resumo

As doenças cardiovasculares são umas das principais doenças que prejudicam a saúde humana, atraindo cada vez mais atenção da comunidade médica. É portanto necessário que os pacientes possam fazer um exame cardiovascular em qualquer instante, sítio e com um resultado preciso no momento. É notar que muitos pacientes com estas doenças são vítimas de ataque cardíaco quando dormem e acabam por morrer por não serem resgatadas a tempo. Baseando este assunto como ponto de partida, esta dissertação foca-se no estudo do sinal do dispositivo SCG ("Seismocardiogram"), o dispositivo registra informações em tempo real do coração através de sensores sem fazer contacto com o corpo, depois utiliza tecnologia de inteligência artificial para analisar as informações registadas e alerta sobre a possível ocorrência de problemas cardíacos.

O sinal da SCG é imperfeição, apesar de ser inicialmente processado por um circuito de hardware, ainda pode conter ruído, com valores incorretos e inconsistentes. Portanto, antes de trabalhar com o sinal, deve-se utilizar a técnica de pré-processamento de sinais para melhorar a qualidade dos sinais por via da eliminação e filtragem. "Wavelet transform" ou Passa banda é um método que pode efetivamente eliminar o ruído.

Outro problema da atualidade, é conseguir identificar os tipos dos problemas cardíacos de forma atempada e eficiente. Devido à necessidade de analisar uma grande quantidade de dados, torna-se difícil conseguir obter estas métricas (identificação e eficiência) e consequentemente ajudar resolver este problema que abrange em grande escala a população mundial. Contudo nos últimos anos, o surgimento de métodos como "Machine Learning" permite fazer uma melhor prevenção sobre a possível ocorrência de problemas cardíacos, tendo-se tornado um das melhores métodos desenvolvidos para tal.

Nesta dissertação pretende-se desenvolver um sistema que recolha dados de uma SCG "Seismocardiogram", que integre um acelerómetro MEMS baseado na medição de tempos de "pull-in", tais como batimentos cardíacos e respiração. Também se pretende desenvolver uma aplicação que utilize "Machine Learning" para reconhecer sinal de SCG. Outro objetivo é utilizar a biblioteca "Open Source" de "TensorFlow" para implementar um algoritmo, que seja capaz de analisar e prever a evolução de cada doente. Desta forma, é possível ter um sistema que, através do histórico de saúde e dos dados recolhidos do doente, possibilite aos médicos detetarem mais facilmente um problema cardíaco que o paciente tenha.

Keywords: "TensorFlow", Patologia cardíaca, "Machine Learning", Prevenção, SCG "Seismocardiogram", Algoritmo

Abstract

Cardiovascular diseases are one of the major threats to human health, attracting more and more attention from the medical community. It is therefore necessary that patients are able to take a cardiovascular examination at any time, place and with a precise result immediately. It is noteworthy that many patients with these diseases are victims of heart attacks when they sleep and end up dying for not being rescued in time. Based on this subject as a starting point, this dissertation focuses on the study of the signal of a SCG device (Seismocardiogram), a device that registers real-time information from the heart through sensors without making contact with the body, after use artificial intelligence technology to analyze the information recorded and warns about the possible occurrence of heart problems.

The SCG is imperfections, and although it is initially processed by an hardware circuit, it may still contain noise, with incorrect and inconsistent values. Therefore, prior to working with the signal, a signal preprocessing technique should be used to improve signal quality through elimination and filtering. Wavelet transform or Passband are methods that can be effectively used to eliminate noise.

Another problem today, is the ability to identify the types of diseases in a timely and efficient manner. Due to the need of analyzing a large amount of data, it is difficult to obtain these metrics (forecasting and efficiency) and consequently to solve this problem that covers the world population on a large scale. However, in recent years, the emergence of methods such as Machine Learning allow for better prevention of the possible occurrence of cardiac problems and has become one of the best methods developed for this.

This thesis intends to develop a system that collects data from an SCG Seismocardiogram which integrates a MEMS accelerometer based on the measurement of pull-in times such as heart rate and respiration. It also intends to develop an application using Machine Learning to recognize an SCG signal. Another goal is to use the TensorFlow Open Text library to implement an algorithm capable of analyzing and predicting the evolution of each patient. In this way, it is possible to have a system that, through the history of health and data collected from the patient, enables physicians to easily detect a cardiac problem that the patient might have.

Keywords: TensorFlow, Cardiovascular Diseases, Machine Learning, Prevention, SCG (Seismocardiogram), Algorithm

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Enquadramento	2
1.3	Objetivos	3
2	Estado da Arte	5
2.1	Problemas de saúde	5
2.2	Doenças cardíacas	6
2.2.1	Enfarte do miocárdio	6
2.2.2	Estenose aórtica	6
2.2.3	Cardiomiopatia dilatada	7
2.2.4	Cardiomiopatia hipertrófica	7
2.3	Exames Cardiológicos	7
2.3.1	Eletrocardiograma	8
2.3.2	Ecocardiograma	8
2.3.3	Sismocardiograma	9
2.3.4	Teste ergométrico	10
2.3.5	Conclusão	10
2.4	Deteção de doenças cardíacas com “Machine Learning”	11
2.5	Algoritmos de “Machine Learning”	13
2.5.1	Árvores de decisão	13
2.5.2	Rede neuronal artificial	14
2.5.3	“Support Vector Machine”	16
2.5.4	Regressão logística múltipla	17
2.5.5	Random forest	18
2.5.6	“Naive Bayes Classifier”	19
2.5.7	“K-Nearest Neighbour”	19
2.5.8	“K-Means”	20
2.5.9	Conclusão	21
2.6	Bibliotecas de “Machine Learning”	22
2.6.1	“TensorFlow”	22
2.6.2	“Apache Mahout”	23

2.6.3	“Scikit-learn”	24
2.6.4	“Accord.NET”	24
2.7	Conclusão do estado da arte	24
3	Análise	26
3.1	Arquitetura do Sistema	26
3.2	Funcionalidades da aplicação	27
3.3	Comportamento do Sistema	28
3.4	Aquisição de Sinais de SCG	28
3.4.1	Filtro de Sinais proveniente do Sensor SCG	29
3.4.1.1	Passa-banda	30
3.4.1.2	Transformada Discreta de Wavelet	30
3.4.2	Identificação de Sinais da SCG	31
3.4.3	Exploração de dados	31
3.4.4	Algoritmos de identificação das sinais da SCG	32
3.4.4.1	“K-Mean clustering”	32
3.4.4.2	Regressão logística múltipla	33
3.4.4.2.1	Função “Softmax”	34
3.4.4.2.2	Função “Cross Entropy”	34
3.4.4.2.3	Função “loss”	35
3.5	Cálculo dos parâmetros da SCG	35
3.5.1	Deteção das ondas AO	36
3.5.1.1	Envelope	37
3.6	Correlação entre os dados do ECO e da SCG	37
3.6.1	Método de correlação de “Pearson”	38
3.7	Método de classificação de doenças cardíacas	39
3.7.1	Diagrama de estados	40
3.7.2	Comparação de ECO e SCG	40
3.7.3	Algoritmos de prevenção de doenças cardíacas	41
3.7.3.1	“Support Vector Machine”	41
3.7.3.1.1	Função “Kernel”	41
3.7.3.1.2	Função “loss”	42
3.7.3.2	Redes neurais artificiais	42

3.7.3.2.1	“Deep Learning”	42
3.7.3.2.2	“Feedforward”	42
4	Especificação do Sistema	44
4.1	Hardware	44
4.1.1	Sensor SCG	44
4.2	Software	45
4.2.1	Linguagem de programação	45
4.2.2	Bibliotecas	45
4.2.2.1	“numpy”	45
4.2.2.2	“scipy”	45
4.2.2.3	“matplotlib”	45
4.2.2.4	“Tensorflow”	46
4.2.2.5	csv	46
4.2.2.6	“seaborn”	46
4.2.2.7	“pandas”	46
4.2.2.8	pywt	46
4.3	Ambiente integrado de desenvolvimento de programas	46
4.3.1	PyCharm	46
4.3.2	Psoc Creator	47
4.4	Dados disponíveis	47
4.4.1	Dados provenientes da SCG	47
4.4.2	Dados provenientes do ECO	47
5	Implementação do Sistema	49
5.1	Fluxograma	49
5.2	Algoritmos de “Machine Learning”	52
5.2.1	“K-Meam clustering”	53
5.2.1.1	Realização dos testes sobre “K-Mean”	53
5.2.2	Regressão logística múltipla	54
5.2.2.1	“softmax” e “loss”	54
5.2.2.2	Realização da avaliação sobre o modelo da regressão logística múltipla	54
5.2.2.3	Realização dos testes sobre o modelo da regressão logística múltipla	55
5.2.3	SVM	55

5.2.3.1	Realização da avaliação sobre o modelo SVM	56
5.2.3.2	Realização dos testes sobre o modelo SVM	56
5.2.4	Rede Neuronal da DNN	56
6	Resultados obtidos	58
6.1	Resultado do sinal recebido	58
6.2	Resultado dos filtros de Sinais	58
6.2.1	Passa-banda	59
6.2.2	Transformada Discreta Wavelet	60
6.2.3	Discussão do resultado	63
6.3	Resultado da identificação de sinais da SCG	65
6.3.1	Resultado do modelo K_Mean	66
6.3.1.1	Resultados dos testes	66
6.3.2	Resultado do modelo regressão logística múltipla	67
6.3.2.1	Resultados dos testes	68
6.3.3	Discussão dos resultados	68
6.4	Análise da SCG	69
6.4.1	Determinação dos picos AO	69
6.4.2	Determinação do sinal da respiração	70
6.4.3	Ondas do sinal da SCG (Ensemble dos Sinais)	70
6.4.4	Análise da correlação entre ECO e ECG	72
6.5	Resultado da identificação da patologia cardíaca	75
6.5.1	Resultados com dados da SCG	75
6.5.2	Resultados com dados do ECO	75
6.5.3	Discussão dos resultados	76
7	Conclusão e trabalho futuro	77
	Referências	79
A	ANEXOS	85
A.1	Correlação entre ECO e SCG	85

Lista de Figuras

1	Divisão do projeto nas duas diferentes partes.	2
2	Mortes globais em 2016	6
3	Pulso típico de ECG e ondas características.	8
4	Imagem de ECO.	9
5	Pulso típico da SCG e ondas características.	9
6	Árvore de decisão.	14
7	Neurónio artificial.	15
8	“Support Vector Machine”.	17
9	Regressão logística múltipla.	18
10	“K-Nearest Neighbour”.	20
11	“K_mean”.	21
12	Funcionalidades gerais do sistema.	26
13	Funcionalidades da aplicação.	27
14	Diagrama do estado do sistema.	28
15	Diagrama do estado da eliminação das sinais inválidos.	29
16	Identificação de sinais da SCG.	31
17	Sistema do cálculo dos parâmetros da SCG.	36
18	Método do cálculo dos parâmetros da SCG.	36
19	Sistema para identificar uma correlação entre ECO e SCG.	38
20	Sistema da determinação de doenças cardíacas.	39
21	Diagrama de estados.	40
22	Dispositivo de medição da SCG.	44
23	Fluxograma do sistema.	49
24	Fluxograma do envio	50
25	Fluxograma da receção	50
26	Fluxograma do sistema.	50
27	Fluxograma da determinação de frequência respiratória	51
28	Fluxograma da determinação de frequência cardíaca	51
29	Fluxograma da determinação da onda médio da sinal da SCG	52
30	Fluxograma da determinação de correlação entre ECO e SCG	52
31	Representação gráfica do sinal obtido pelo dispositivo da SCG.	58

32	Sinal proveniente do paciente DCM	59
33	Sinal proveniente do paciente HCM	59
34	Sinal proveniente do paciente MI	59
35	Sinal proveniente do paciente AS	60
36	Sinal proveniente do paciente Healthy	60
37	Resultados da decomposição	61
38	Sinal proveniente do paciente DCM	62
39	Sinal proveniente do paciente HCM	62
40	Sinal proveniente do paciente MI	62
41	Sinal proveniente do paciente AS	62
42	Sinal proveniente do paciente Healthy	63
43	Sinal proveniente do paciente DCM	63
44	Sinal proveniente do paciente Healthy	64
45	Sinal proveniente do paciente MI	64
46	Sinal proveniente do paciente HCM	64
47	Sinal proveniente do paciente AS	65
48	Resultado do modelo K-Mean	66
49	Teste sucedido	67
50	Teste sucedido	67
51	Teste sucedido	67
52	Teste sucedido	67
53	Teste sucedido	67
54	Teste falhado	67
55	Resultado do modelo regressão logística múltipla.	68
56	Teste falhado	68
57	Teste falhado	68
58	Análise do sinal da SCG adquirido e determinação dos picos AO	69
59	Sinal correspondente ao ciclo da respiração e determinação da frequência respiratória	70
60	Ensemble dos sinais da SCG adquiridos.	71
61	Parâmetros calculados	71
62	A correlação entre End-diastolic diameter e IM_MO	73
63	A correlação entre Ejection Fraction e IM	73
64	A correlação entre Stroke Volume e AO	74

65	A correlação entre Ejection Time e AO_ACM	74
66	Correlação entre SV e ondas de SCG	85
67	Correlação entre PEP e ondas de SCG	85
68	Correlação entre EF e ondas de SCG	86
69	Correlação entre IVSd e ondas de SCG	86
70	Correlação entre PP e ondas de SCG	87
71	Correlação entre LVMass e ondas de SCG	87
72	Correlação entre LVM e ondas de SCG	88
73	Correlação entre LVEd e ondas de SCG	88
74	Correlação entre EF e ondas de SCG	88
75	Correlação entre MFR e ondas de SCG	89

Lista de Tabelas

1	Regressão logística múltipla vs K-Mean	68
2	DNN vs SMV baseados na base de dados da SCG	75
3	DNN vs SMV baseados na base de dados do ECO	76

Lista de Equações

1	Equação do cálculo da distância	33
2	Equação do cálculo da centroide	33
3	Equação de “Softmax”	34
4	Equação de “Cross_Entropy”	34
5	Equação de “loss”	35
6	Equação de RBF	41
7	Equação de “loss”	42

List of Algorithms

1	“K-Mean”	53
2	Teste	53
3	Regressão logística múltipla	54
4	“Softmax”	54
5	“Accuracy”	55
6	Teste	55
7	SVM	55
8	RBF	56
9	“Accuracy”	56
10	Teste	56
11	Redes neuronais da DNN	57

Lista de Abreviaturas

ECG	Eletrcardiograma
SCG	Seismocardiogram
ECO	Ecocardiograma
OMS	Organização Mundial da Saúde
MI	Myocardial Infarction
AS	Aortic Stenosis
DCM	Dilated Cardiomyopathy
HCM	Hypertrophic Cardiomyopathy
SVM	Support Vector Machine
CPU	Central Processing unit
GPU	Graphics Processing unit
TPU	Tensor Processing unit
API	Application programming interface
BLE	Bluetooth Low Energy
DC	Débito Cardíaco
AS	Atrial systole
IM	Isovolumic Moment
AO	Aortic Opening
IC	Isovolumic Contraction
RE	Rapid Ejection
AC	Aortic Closure
MO	Miral Opening
HR	Heartrate
SV	Systolic Volume
PEP	Pre-Ejection period
EF	Ejection Fraction
IVSd	Interventricular sepetal thinckness
LVMass	Left Ventricle Mass
EDV	End-Diastolic Volume
LVED	End-Diastolic Diameter
ET	Ejection Time

MFR Maximum Flow Rate

DNN Deep Neural Networks

RBF Radial Basis Function

1 Introdução

1.1 Motivação

Hoje em dia as doenças cardíacas são um dos problemas de saúde mais graves, afetando cerca de 23 milhões de pessoas em todo o mundo. Entre um a dois por cento da população adulta nos países desenvolvidos tem problemas cardíacos, sendo também estimado que 10% do grupo etário com idades superiores a 70 anos sofram deste problema. De acordo com a Organização Mundial da Saúde (OMS), cerca de 17,5 milhões de pessoas morrem todos os anos pelas doenças cardíacas sendo esta a principal causa de morte em todo o mundo. A OMS mostra ainda que mais de 75% das mortes provocadas por doenças cardiovasculares são registadas em países pouco desenvolvidos, sendo que 80% dos óbitos são causados especificamente por ataques cardíacos [1].

O sistema cardiovascular tem como principal função impulsionar o sangue para todo o organismo. Hoje em dia, os maus hábitos, como a má alimentação, stresse, tabaco e álcool, bem como fatores genéticos, proporcionam um aumento do risco de desenvolvimento de doenças cardíacas [2]. Sendo que estas doenças são mais comuns em pessoas mais idosas e com histórico familiar. Contudo, isso não significa que pessoas mais jovens e sem histórico familiar de doenças cardíacas sejam imunes.

As doenças cardíacas podem afetar o coração de várias formas, provocando paragens cardíacas, alterações nos batimentos cardíacos e disfunções da aorta. Dentro dos principais problemas cardíacos encontram-se a cardiomiopatia hipertrófica, popularmente conhecida como doença do músculo cardíaco, que se caracteriza pelo engrossamento do músculo do coração. Isto faz com que seja mais difícil para o coração bombear o sangue e consequentemente provoca uma deficiência funcional do músculo cardíaco. Outro problema cardíaco muito comum é o enfarte, que se caracteriza por um coágulo que impede a passagem do sangue em parte da câmara cardíaca [3].

Atendendo ao elevado número de casos registados de doenças cardíacas a nível mundial, e à gravidade dos problemas que daí advêm, torna-se necessário apostar na prevenção atempada destas patologias. Recorrendo às recentes tecnologias de “Data Mining” e “Machine Learning”, poderá ser possível analisar dados de pacientes e tentar prever problemas cardíacos.

Apesar de as tecnologias de “Data Mining” e “Machine Learning” serem frequentemente adotadas para prever problemas cardíacos, os seus desempenhos são geralmente afetados pelo estado dos dados. Diferentes conjuntos de dados podem apresentar diferentes características, dimensões ou formato, pois, os dados podem ser numéricos ou simbólicos, conter ruído ou imperfeições, incorretos, inconsistentes, duplicados e ausentes.

As técnicas de pré-processamento de dados são frequentemente utilizadas para melhorar a qualidade

dos dados por via eliminação ou minimização dos problemas citados. Esta melhoria pode facilitar a utilização de técnicas de “Data Mining” e “Machine Learning”.

1.2 Enquadramento

O projeto no qual esta dissertação se encontra inserida aborda o problema descrito anteriormente, através da implementação de algoritmos de “Machine Learning”, utilizando a biblioteca “TensorFlow” para detetar patologias cardíacas. Através destes algoritmos, poderá ser possível analisar a evolução de cada doente e facilitar ao médico a deteção de doenças cardíacas. Este tipo de sistema está evoluindo de modo crescente na cardiologia e já existem excelentes exemplos em diversas áreas, tal como um sistema para interpretação eletrocardiográfica que conseguiu reconhecer automaticamente padrões eletrocardiográficos com precisão de 88% para classificação dos ritmos anormais e um sistema para interpretação de eco cardiográfica que também é muito estudada e foi capaz de assistir com excelente precisão. Em termos da SCG, esta não é tão conhecida como os dois exemplos anteriores, mas também é capaz de avaliar de forma discreta os sinais cardíacos, sem necessidade de fazer contacto com o corpo, assim tornando-se numa vantagem para monitorização cardíaca em situação da vida quotidiana.

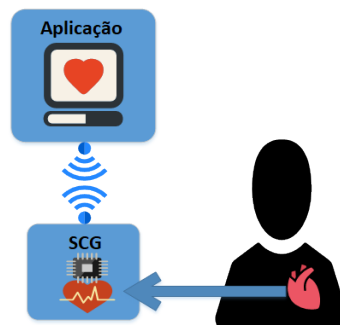


Figura 1: Divisão do projeto nas duas diferentes partes.

Como se pode verificar pela Figura 1, o sistema será composto por duas partes: sistema embebido integrado com o aparelho SCG e aplicação incorporada com algoritmos de “Machine Learning”. O sistema embebido integrado com SCG tem como função principal recolher sinais da SCG e serem transmitidos para a aplicação. A aplicação tem como função organizar dados de doentes e armazená-los. Por último, o algoritmo de “Machine Learning” vai ser implementado na aplicação, tendo como função analisar o progresso de cada doente baseando nos dados transmitidos. Esta análise tem como função a verificação de que a informação obtida se encontra nos padrões normais, ou se os dados se distanciam dos valores referenciados, sendo assim necessário informar, de forma atempada, os médicos.

Desta forma, estas duas partes serão desenvolvidas durante a dissertação e tem como objetivo de implementar um algoritmo mais eficaz para prever as patologias cardíacas.

1.3 Objetivos

O principal objetivo deste trabalho consiste em definir um sistema para determinar ou prever os tipos de doenças cardíacas, para diferentes pacientes, dependendo do funcionamento do seu coração.

Para conseguir atingir este objetivo principal será necessário o cumprimento dos seguintes objetivos.

- Desenvolver um sistema embebido que recolha sinais de SCG;
- Definir um modelo de “Machine Learning” para identificar sinais válidos e inválidos da SCG;
- Implementar um programa para extrair as características ou informações necessárias a partir de sinais tais como o batimento cardíaco e a respiração;
- Identificar uma correlação entre os dados de SCG e os dados de ECO;
- Definir um modelo de “Machine Learning” para detetar doenças cardíacas;

2 Estado da Arte

Este capítulo está organizado em seis secções principais. Inicialmente apresenta-se o problema de saúde na sociedade e mostra que a maior risco de saúde é a doença cardíaca. De seguida, pretende-se apresentar algum típico de doenças cardíaca. Depois, pretende-se ilustrar os dados proveniente de diversos métodos diagnósticos que possam ser utilizados para prevenir o risco de doenças cardíacas. Para além disso, são apresentados também as aplicações de “Machine Learning” para o tratamento de doenças cardíacas e diversos algoritmos com maior precisão da deteção de doenças cardíacas. Em fim, são mostradas as diversas bibliotecas de “Machine Learning” que podem ser utilizadas no mercado, reconhecendo as vantagens e desvantagens destas técnicas.

2.1 Problemas de saúde

Uma das características da sociedade atual é a grande quantidade de pessoas que atinge idade avançada, o que torna um problema novo a ser enfrentado pela ciência. Com o progresso mundial, o aumento cada vez número de pessoas prepara-se para uma vida mais saudável, com melhores perspectivas de vida social.

O número de pessoa idosa vem crescendo gradativamente nos últimos anos, devido às melhorias de condições de vida e dos avanços da tecnologia. Entretanto, o processo de envelhecimento acompanha se de perdas funcionais, que facilitam o aparecimento de doenças, piorando a qualidade de vida, especialmente o funcionamento cardíaco. A saúde do coração é importantíssima e felizmente cada vez mais pessoas lhe dão a devida atenção. O coração é um dos principais órgãos do corpo humano que bombeia o sangue através dos vasos sanguíneos do sistema circulatório. O sangue transporta oxigénio e nutrientes para corpo. Se o coração não funcionar corretamente, isso levará sérios problemas de saúde. Existem diversos tipos de problema do coração, os quais podem ser agrupados em imutáveis e mutáveis. Os tipos imutáveis são aqueles que não podem mudar e tratar: idade, hereditariedade e outros. Os tipos mutáveis são os quais podem controlar e prever: fumo, pressão arterial e outros.

De acordo com dados da organização Mundial da Saúde e , mostram que a maior causa de mortalidade e morbidade é a doença cardiovascular que isso é problema do coração ou dos vasos sanguíneos, que predispõe uma pessoa a maior risco de saúde [4]. E a Organização Mundial de Saúde conclui que o ataque cardíaco é a causa de aproximadamente 31% de todas as mortes ao redor do globo, tanto em homens como em mulheres e a insuficiência cardíacas congestiva, mais comum de morbidade e mortalidade na população idosa [5]. Em Portugal, cerca de 35 mil portugueses morrem anualmente por doenças cardiovasculares, que apresentam um terço de toda a mortalidade da população, e é a primeira causa de morte [6].

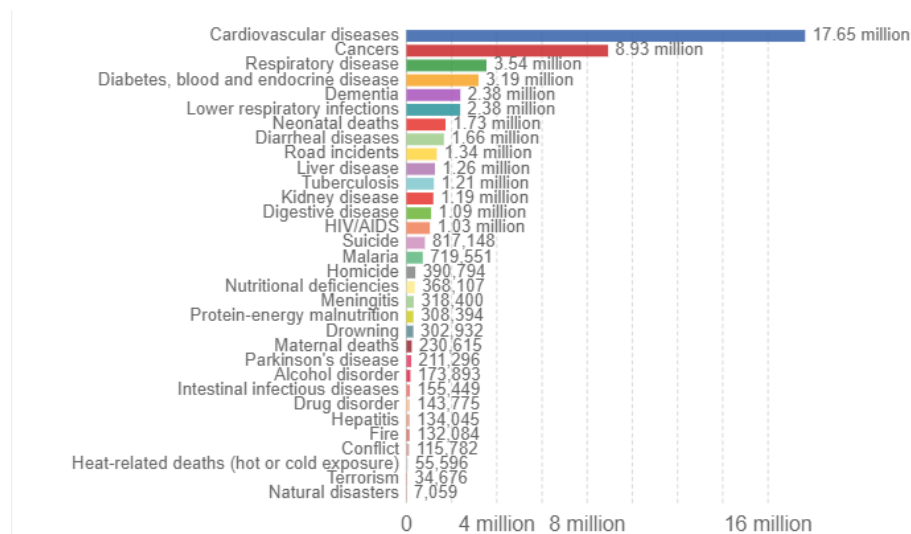


Figura 2: Mortes globais em 2016

Na Figura 2 mostra as mortes globais por causa em 2016, ordenadas de maior número de mortes para menor. Como o gráfico mostrado, em 2016, cerca de 55 milhões de pessoas morreram e o principal causa de mortes foi doença cardiovasculares, pois esta doença foi responsável pela morte de um terço de todas as mortes que ocorreram em 2016 (cerca de 17.6 milhões de pessoas) [7].

2.2 Doenças cardíacas

Sem tratamento rápido, um ataque cardíaco pode levar a sérios problemas de saúde e até a morte. Conforme já mencionado, o ataque cardíaco é uma coisa comum de morte em todo o mundo. Alguns dos sintomas comuns de ataque cardíaco são os seguintes.

2.2.1 Enfarte do miocárdio

Enfarte do miocárdio ou “Myocardial infarction” (MI) é a designação dada um “ataque cardíaco” que provoca a morte do tecido muscular cardíaco pelo entupimento de uma artéria coronárias que é o vaso responsável pela chegada de oxigênio e outros nutrientes ao músculo cardíaco. O enfarte é uma emergência média que requer tratamento imediato.

Em cada ano, o Enfarte mata, em média, cerca de 4342 pessoas, das quais 2420 homens e 1922 mulheres, o que revelem que são cerca de 360 pessoas que morrem, por mês, desta doença [8].

2.2.2 Estenose aórtica

Estenose Aórtica ou “Aortic Stenosis” (AS) apresenta um problema da válvula aórtica, que é a componente do coração responsável pela passagem de sangue para o corpo. Este problema faz com que a válvula

aórtica fique progressivamente mais rígida, o que dificulta a ejeção do sangue bombeado pelo coração, criando assim um obstáculo à saída do sangue do coração para o corpo. Esta doença é rapidamente fatal e incapacitante, sendo sobretudo uma doença de desgaste associada a pessoas idosas.

Tendo em conta um estudo realizado no dia mundial dos Avós, em Portugal, mais de 30 mil idosos sofrem de estenose aórtica, o que indica que um em cada 15 portugueses com mais de 80 anos tem esta doença [9].

2.2.3 Cardiomiopatia dilatada

Cardiomiopatia dilatada ou “Dilated CardioMyopathy” (DCM) é causada por alteração do músculo cardíaco em que os ventrículos aumentam de tamanho, o que provoca incapacidade de bombear sangue suficiente para satisfazer as necessidades do corpo, tornando-se em insuficiência cardíaca. Este problema pode surgir em qualquer idade, mas é mais frequente em pessoas com idades entre os 20 e 60 anos, tendo mais incidência em indivíduos do sexo masculino.

Anualmente, cerca de 10% das pessoas que apresentam este problema têm mais de 65 anos, o que significa que cerca de cinco a oito pessoas em 100.000 apresentam esta doença [10].

2.2.4 Cardiomiopatia hipertrófica

Cardiomiopatia hipertrófica ou “hypertrophic cardiomyopathy” (HCM) é uma doença do músculo cardíaco que é caracterizada pelo engrossamento do músculo do coração, criando uma deficiência funcional do músculo cardíaco. Esta doença é o tipo mais comum de doença cardíaca de origem genética, ou seja, a maioria das pessoas com este problema têm-no devido a um defeito genético hereditário.

De acordo com o prognóstico, cerca de 1 a 3% de adultos com cardiomiopatia hipertrófica morrem a cada ano. Crianças com este problema têm mais tendência a morrerem [11].

2.3 Exames Cardiológicos

Embora todas estas doenças sejam preocupantes já existem diversos diagnósticos, que qualquer pessoa pode fazer, de maneira a preveni-las ou a diminuir o risco de as desenvolver. Estes métodos servem para avaliar o funcionamento do coração, e algumas vezes fornecem informação sobre a estrutura do coração e possíveis doenças cardíacas. Além disso, é possível ver o ritmo, frequência do batimento cardíaco e outras características do coração. Atualmente, os diagnósticos de cardiologia evoluíram e multiplicaram-se, sendo que os principais métodos que detetam possíveis doenças do coração são: Eletrocardiograma; Ecocardiograma; Teste ergométrico; e Sismocardiograma.

2.3.1 Eletrocardiograma

Eletrocardiograma (ECG) é um dos exames mais utilizados pelos cardiologistas. É um sistema rápido, simples e indolor no qual impulsos elétricos do coração são amplificados, detetados e registados. O sistema fornece informações sobre a parte do coração que provoca cada batimento cardíaco, cada contração do músculo cardíaco e cada movimento das válvulas do coração. Por vezes, os dados da eletrocardiograma podem mostrar um aumento do músculo cardíaco, devido a uma deficiência funcional ou que a coração não está a receber oxigênio suficiente devido a um dos vasos sanguíneos ter ficado progressivamente rígido.

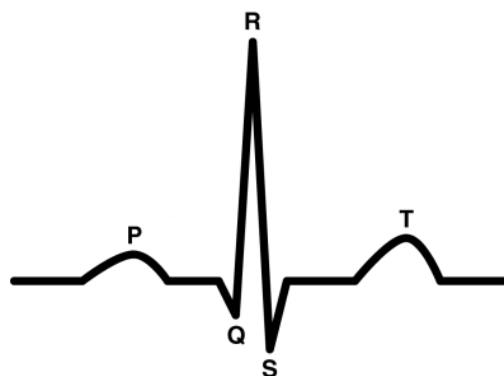


Figura 3: Pulso típico de ECG e ondas características.

A Figura 3 ilustra uma representação do comportamento do traçado eletrocardiográfico, sendo este um gráfico da variação de tensão em relação ao tempo. As variações resultam do movimento do músculo cardíaco. O gráfico é constituído pelas ondas P, Q, R, S e T, cada uma delas tendo o seu significado relacionado com os funcionamentos de coração. A onda P representa o funcionamento atrial, e a sua duração pode ser causada pela Hipertrofia atrial. As ondas QRS correspondem ao funcionamento ventricular, são maiores que a onda P pois a massa muscular dos ventrículos é maior do que a dos átrios. A onda T caracteriza o músculo ventricular em repouso, a sua configuração anormal indica hipercalemia.

A eletrocardiograma é muito aplicada quando há suspeitas de arritmia, taquicardia ou bradicardia. Às vezes, também é apropriada para detetar o fluxo inadequado de sangue para o músculo cardíaco. Além disso, ele pode ser utilizado como base de comparação com informações posteriores.

2.3.2 Ecocardiograma

O Ecocardiograma (ECO) é um exame de imagem feito usando ultrassonografia. É um sistema indolor, rápido e inócuo, podendo ser utilizado sem qualquer risco sempre que necessário. O sistema é capaz de diagnosticar a estrutura interna do coração através das ondas de som por ele emitidas e permite obter informação morfológica e funcional sobre o músculo cardíaco, as válvulas cardíacas e os vasos sanguíneos.

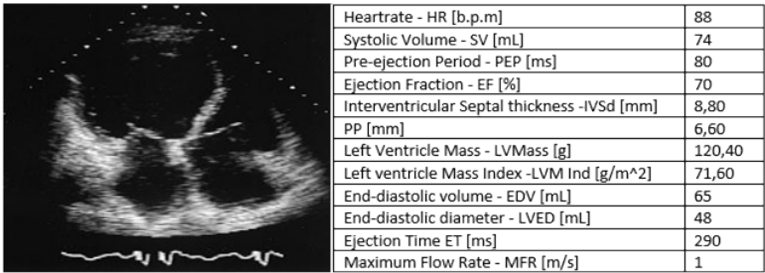


Figura 4: Imagem de ECO.

Como a Figura 4 mostra a interpretação dos resultados da ecocardiografia é mais acessível que a de eletrocardiografia, pois Ecocardiogramas são exames que produzem imagens estáticas e o movimento do coração e dos vasos sanguíneos que lhe estão próximos em tempo real. Além disso, através da imagem, é possível identificar a direção, a velocidade do fluxo sanguíneo no interior das cavidades cardíacas, o volume de sangue bombeado pelo ventrículo, a espessura da parede do ventrículo e o diâmetro diastólico.

O Ecocardiograma é indicado sempre que se suspeita da presença de lesões nas válvulas cardíacas, nas paredes do coração ou de um inadequado processo de bombeamento do sangue. É destinado para pessoas que sofrem de cardiomiopatia hipertrófica, insuficiência cardíaca, enfarte, problema nas válvulas, anomalias congênitas entre outras. Os resultados são normalmente mais precisos do que os obtidos num raio X, além de não expor o paciente à radiação.

2.3.3 Sismocardiograma

Sismocardiograma (SCG) é um sistema que convencionou um método não invasivo que registra a vibração de baixa frequência causada pelo movimento do coração, utilizando um acelerómetro MEMS. Para a medição, através desse sistema, não há necessidade de um contacto direto com o utilizador. O sistema é capaz de recolher bastante informação para determinar o comportamento cardíaco em vários aspetos e pode ser utilizado sem a presença de médicos ou especialistas, pois não depende do operador durante a aquisição dos dados.

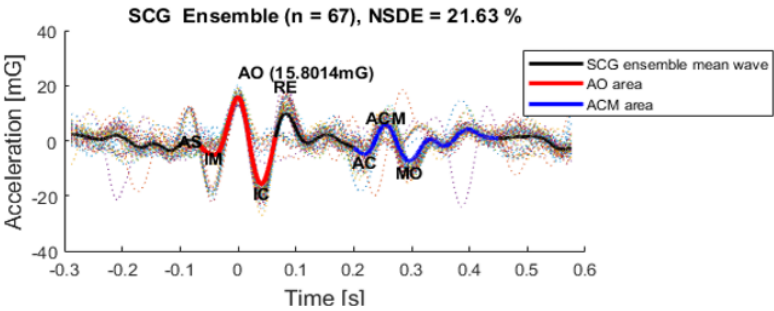


Figura 5: Pulso típico da SCG e ondas características.

Na Figura 5 está presente um gráfico da variação da aceleração em relação ao tempo. O gráfico é constituído pelas ondas AS, IM, AO, IC, RE, AC, ACM e MO, e cada uma delas está relacionada a eventos cardíacos significativos. A onda AS descreve o funcionamento do sístole atrial. A onda IM descreve o momento Isovolumétrico durante o ciclo sistólico. A onda AO revela a abertura da válvula aórtica. A onda IC representa a contração isovolumétrica. A onda RE retrata o enchimento rápido, que é progressivo nos ventrículos e o enchimento dos átrios, a pressão dentro destes últimos supera a pressão dentro da cavidade ventricular e as valvas atrioventriculares abrem-se. A onda AC caracteriza o fecho da válvula aórtica durante o ciclo diastólico. ACM representa o período diastólico. Por último, a onda MO representa o funcionamento da abertura da válvula mitral. Para além disto, a combinação entre as ondas pode fornecer outro tipo de informação útil sobre o comportamento do coração, como o intervalo de tempo entre AO e AC que pode corresponder ao tempo de ejeção sistólica e o intervalo de tempo entre MC e AO que pode corresponder ao tempo de contração isovolumétrica [12].

Este exame não é muito conhecido como os dois exames anteriores, mas baseia-se na informação recolhida também é possível prever ou avaliar algumas patologias cardíacas como o enfarte, a cardiomiopatia hipertrófica, a cardiomiopatia dilatada e a estenose Aórtica.

2.3.4 Teste ergométrico

o teste ergométrico é também conhecido como teste de Esforço, é um exame não invasivo que consiste na aplicação de um esforço físico com incremento progressivo na intensidade do exercício. Este exame é acompanhado por um eletrocardiograma que mede o batimento do coração do paciente. O sistema regista o ritmo cardíaco deste momento de esforço de baixa intensidade, até às possíveis alterações que podem ocorrer durante o ritmo acelerado.

Este exame é normalmente usado para a prevenção de entupimento das artérias do coração e é indicado para detetar arritmias, problemas nas válvulas cardíacas, doenças congénitas e alterações do ritmo cardíaco. Apesar de ter a capacidade de fornecer informações muito relevantes, esta exame não deve ser realizado por pacientes que têm limitações físicas, como impossibilidade de andar ou pedalar, ou que possuem alguma doença cardíaca, como enfarte do miocárdio ou insuficiência cardíaca. Com este exame é possível avaliar imediatamente fenómenos anormais cardíacos durante o exame.

2.3.5 Conclusão

Relativamente aos exames enunciados, todos eles são eficazes, para o diagnóstico da patologia cardíaca, no entanto, cada um deles apresenta algumas características especiais. O eletrocardiograma é um exame barato, de fácil interpretação e fornece um grande número de informações sobre o coração, mas apresenta

baixa sensibilidade em detetar alguma doenças cardíacas [13]. O ecocardiograma comparado ao exame anterior já possuiu uma boa sensibilidade e permite a avaliação dinâmica das massas intracardíacas com boa precisão, comparado a outros exames, mas apresenta a reprodutibilidade deficiente e alto custo [14]. O sismocardiograma tem a capacidade de recolher bastante informações e determinar o comportamento cardíaco em vários aspetos, mas devido ao muito complexo sinal do sismocardiograma, a sua interpretação torna-se difícil. O teste ergométrico, às vezes, é um teste preciso, porém a sua realização não é aconselhada para pacientes com limitações físicas [15].

2.4 Detecção de doenças cardíacas com “Machine Learning”

A evolução tecnológica está a melhorar o mundo, sobretudo, na área da saúde, onde houve um grande avanço nas tecnologias médicas. A tendência do momento é “Machine Learning”, que tem oferecido grandes soluções e inovações para facilitar o trabalho dos médicos, ajudando-os a reduzir os erros possíveis e alcançando um maior grau de precisão. “Machine Learning” é exatamente o conceito segundo o qual as “máquinas aprendem como os humanos.”, um técnica desenvolvida para ser capaz de criar, de forma autónoma e a partir de experiências anteriores, uma hipótese, ou função, capaz de dar uma resposta ao problema que deseja resolver. A capacidade de aprendizagem é considerada essencial para esta técnica. Atividades como memorizar, observar e explorar situações para aprender fatos, melhorar resultados através da prática e organizar conhecimentos novos.

Os métodos de “Machine Learning” podem ser classificados em 3 categorias: Aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço. A aprendizagem supervisionada supõe uma supervisão no conjunto de treino que indica a resposta pretendida para cada entrada, isso é, dado um conjunto de dados na entrada, já se sabe como deve ser saída correta, ou seja, existe sempre uma relação entre a entrada e saída. Esta categoria é normalmente aplicada em problemas de regressão e classificação. Num problema de regressão, tenta-se prever os resultados numa saída contínua, ou seja, tenta-se mapear as entradas para uma função contínua. Num problema de classificação, procura-se os resultados numa saída discreta, isso é, tenta-se mapear as entradas em categorias distintas. A aprendizagem não supervisionada é o oposto da categoria anterior, permite resolver problemas com pouca ou nenhuma indicação, nesta categoria não há feedback com base nos resultados da previsão, ou melhor, as informações de classes oferecidas são desconhecidas. Este tipo de aprendizagem é normalmente utilizada em problemas de agrupamento. Num problema de agrupamento, tenta-se agrupar os dados com base na relação entre as variáveis e os dados. A aprendizagem por reforço é a interação com o meio ambiente e assim descobrir qual a melhor resposta para cada entrada, e podendo ser aplicada em varias áreas.

Há alguns anos atrás, a área de “Machine Learning” era vista como uma área teórica, com aplicações

apenas em pequenos problemas, desafiantes, mas de pouco valor prático. A partir da década de 1970, verificou-se uma maior disseminação do uso de técnica de “Machine Learning” para a resolução de problemas reais. Muitas vezes, estes problemas eram computacionalmente tratados por um sistema baseado no conhecimento e um sistema de decisão. Uma das principais utilizações de “Machine Learning” é na área da saúde para ajudar nos diagnósticos de diversas doenças, o sistema identifica padrões entre milhões de dados de diferentes características e prevê o comportamento por meio de algoritmos com capacidade de aprender e evoluir com base no histórico [16]. As análises preditivas resultadas por este sistema permitem classificar determinadas doenças. Nestas duas últimas décadas, vários estudos têm tido como foco diagnósticos de doenças cardíacas e esses estudos envolvem uma variedade de algoritmos de “Machine Learning” para previsão e alcançaram altas precisões nos resultados.

Muitos dos exemplos destes estudos foram feitos em alguns anos atrás. Em 1989, “Genari” alcançou uma precisão de 78.9% no conjunto de dados da doença cardíaca de Cleveland, utilizando o sistema de agrupamento conceitual. Em 1987, “Detrano”, “Steinbrunn” e “Pfisterer” utilizaram a função discriminante derivada da regressão logística para classificar doenças cardíacas com uma precisão de aproximadamente 77%. Em 2005, “Yao”, “Lei” e “Yin” utilizaram um novo modelo baseado em “Decision Tree”, que melhorou a eficiência da detecção da doenças cardíacas. Em 2008, “Sengur” e “Turkoglu” inventaram uma metodologia, utilizando “Machine Learning” com algoritmos de redes neurais artificiais para prever doenças cardíacas. No mesmo ano, “Palaniappan” e “Awang” desenvolveram um sistema inteligente de previsão de doenças cardíacas, utilizando algoritmos de “Machine Learning”, tais como “Decision Tree”, rede neurais artificial e “Naïve Bayes”. Em 2009, “Shantakumar” e “Kumaraswamy” propuseram um método inovador, utilizaram o algoritmo de “K-means” em combinação com o algoritmo “Maximal Frequent Itemset Algorithm” (MAFIA) para prever doenças cardíacas. Em 2010, “Bala” e “Chana” validaram a precisão dos modelos de previsão usando abordagens de “Machine Learning” em ambiente de “Cloud”. Em 2011, “Chowdhury”, “Chatterjee” e “Samanta” melhoraram a precisão de 75% ao exibir uma previsão baseada na rede neurais artificial [17].

De acordo com a evolução da história, os investigadores esforçam-se cada vez mais nos estudos de pesquisa das doenças cardíacas. Esses estudos empregam uma variedade de algoritmos de “Machine Learning” para previsão e detecção dos problemas cardíacos, com o objetivo de alcançar melhores valores de precisão. Consoante os resultados obtidos, a técnica de “Machine Learning” mostrou resultados favoráveis, o que significa que é uma técnica eficaz na previsão e detecção de doenças cardíacas. Os resultados obtidos também sugerem que a técnica é capaz de melhorar a prestação dos cuidados de saúde, auxiliando e ajudando os médicos a avaliar o risco dos pacientes.

2.5 Algoritmos de “Machine Learning”

A técnica de “Machine Learning” pode inaugurar uma nova era nos cuidados de saúde e ganhou cada vez mais popularidade nos últimos anos. Como o “Big Data” é a tendência em voga em várias áreas do momento, a utilização de “Machine Learning” é poderoso para fazer previsões ou sugestões com base em grandes quantidades de dados. Assim sendo, pretende-se mostrar uma extensa pesquisa de métodos de “Machine Learning” que são os mais frequentemente utilizados na área da previsão e detecção de doenças cardíacas usando “data mining”.

Antes de mais, pretende-se apresentar dois fatores importantes que tem impacto direto em “Machine Learning”: “accuracy” e precisão. A “accuracy” é a mais comum e simplesmente mede com que frequência o modelo de “Machine Learning” faz a previsão correta durante treino ou aprendizagem. É a razão entre o número de previsões corretas e o número total de previsões. A precisão é definida como o número de verdadeiros positivos dividido pelo número de verdadeiros positivos mais o número de falsos positivos. Os falsos positivos são casos em que o modelo classifica incorretamente positivos que são negativos e os verdadeiros positivos são casos em que modelo identifica corretamente os positivos.

2.5.1 Árvores de decisão

As árvores de decisão utilizam uma estratégia, dividir-para-conquistar, decompondo um problema complexo em subproblemas mais simples, aos quais recursivamente é aplicada a mesma estratégia. As soluções dos subproblemas podem ser combinadas formando uma árvore para produzir uma solução do problema primário. A construção de uma árvore de decisão baseia-se na escolha de um atributo que servirá para particionar um conjunto de dados em subconjuntos que correspondem a cada valor do atributo. Assim sendo, é realizada uma observação para verificar se esses subconjuntos pertencem à mesma classe. Se isso acontecer, associa-se essa classe a uma folha contendo todos os dados desse subconjunto. Caso contrário, é escolhido um novo atributo para particionar os dados e o processo é repetido até que haja uma homogeneidade das classes de um determinado subconjunto. Porém, uma árvore pode crescer muito, tornando-se específica para uma determinada base de dados, causando assim um overfitting.

As árvores de decisão são apresentadas como um algoritmo de classificação, em que utiliza uma base de dados para construir árvores de decisão. Uma árvore de decisão é um grafo acíclico direcionado em que cada nó ou nó de divisão, com dois ou mais sucessores ou folhas. Uma folha é rotulada com uma função. Um nó de divisão contém um teste condicional baseado nos valores de entrada ou do atributo. Na Figura 6 está representada uma árvore de decisão e a divisão correspondente no espaço definido pelos atributos x_1 e x_2 . Cada nó da árvore corresponde a uma região desejada. As regiões definidas pelas folhas da árvore

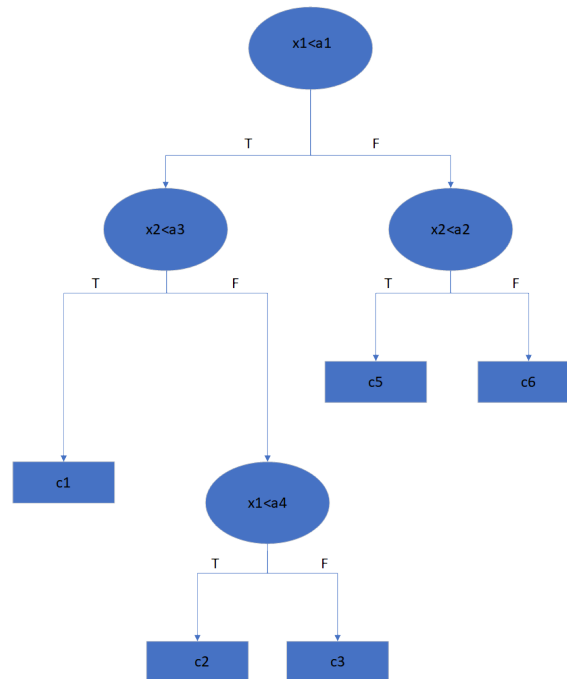


Figura 6: Árvore de decisão.

são mutuamente únicas, e a regiões dessas regiões cobrem todo o espaço definido pelo atributo.

As árvores de decisão são uma das abordagens do método preditiva usada pela técnica de “data mining” e normalmente utilizada para problemas de classificação e previsão. Este algoritmo tem sido aplicado em várias áreas, pois é fácil de entender e interpretar, tendo também a capacidade de lidar com valores ausentes, é também poderosa e robusta.

Este algoritmo é uma das “Data mining” mais bem-sucedidas utilizadas no diagnóstico de cardiopatia. No entanto, a sua precisão não é perfeita [18] [19]. Conforme alguns estudos realizados, a maior precisão alcançada é de 84,1% [18].

2.5.2 Rede neuronal artificial

A rede neuronal artificial tem inspiração nas redes neurais biológicas presentes no cérebro. É um sistema computacional distribuído e composto de unidades por processamento simples, densamente interconectadas. Estas unidades são conhecidas como neurónios artificiais e apresentam um modelo matemático. São também distribuídas numa ou mais camadas e conectadas por um grande número de conexões, normalmente unidirecionais. Estas conexões possuem um peso associado, que ponderam a entrada recebida por cada unidade. Os pesos podem ter um valor positivo ou negativo, consoante o comportamento da conexão, que pode ser excitatório ou inibitório. O valor do peso é ajustado por um processo de aprendizagem e conhecimento adquirido pela rede.

Este algoritmo é, portanto, caracterizado por dois aspectos principais: a arquitetura e processo de aprendizagem. Enquanto a arquitetura está relacionada com o tipo, o número de unidades de processamento e com a forma como os neurónios estão interligados, o processo da aprendizagem diz apenas respeito às regras utilizadas para o ajuste dos pesos e informação utilizadas pelas regras. Como a Figura 7.

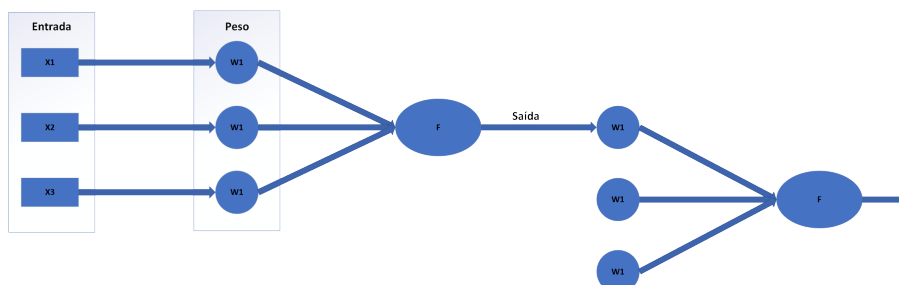


Figura 7: Neurónio artificial.

O neurónio é a unidade de processamento fundamental de uma rede neural artificial. Na Figura 7 é apresentado um modelo simples de um neurónio artificial. Cada terminal de entrada do neurónio, simulando os dendritos, recebe um valor do peso. Os valores recebidos são ponderados e combinados por uma função matemática f , equivalente ao processamento realizado pela soma. A saída da função é a resposta produzida pelo neurónio para a entrada de próximo neurónio.

Conforme já mencionado, numa rede neuronal artificial, os neurónios podem estar dispostos numa ou mais camadas. Quando mais camadas são utilizadas, um neurónio pode receber nos seus terminais de entrada valores de saída de neurónios da camada anterior ou enviar o seu valor de saída para o terminal de entrada dos próximos neurónios. Uma rede com mais do que uma camada de neurónios tem o nome como de rede multicamadas. A camada de um neurónio que produz os valores de saída é denominada de camada de saída. As outras camadas são denominadas de camadas intermediárias ou escondidas. Numa rede multicamadas, as conexões entre os neurónios podem ser classificadas como conexões completas, conexões parciais ou conexões locais. As Conexões completas são quando os neurónios da rede estão ligados a todos os neurónios da camada anterior ou/e seguinte. As conexões parciais são quando os neurónios estão ligados a apenas alguns dos neurónios da camada anterior ou/e seguinte. As conexões locais são quando os neurónios estão ligados a um neurónio que se encontra numa região bem definida.

Como qualquer outro método de “Machine Learning”, as redes neuronais artificiais são também um sistema que aprende a partir dos dados. Elas têm sido usadas para resolver uma grande variedade de tarefas que são difíceis de resolver pelo humano. Como tem com base a estrutura do sistema nervoso humano, têm muitas vantagens em relação a outros métodos. O que leva ao constante aumento do número de aplicações que incorporam este algoritmo confirmando a sua importância para a resolução de problemas práticos.

Devido à ampla gama de aplicabilidade da rede neuronal e à sua capacidade de aprender relacionamentos complexos, incluindo informações ruidosas ou menos precisas, as redes neurais são adequadas para resolver problemas na área biomédica. A técnica de rede neuronal é adotada para classificação de dados médicos, ajuda os especialistas do domínio, fornece ao paciente resultados de diagnóstico precoce, uma vez que ele apresenta um desempenho realista mesmo sem reciclar. Segundo alguns estudos este algoritmo é confiável para diagnósticos de doenças cardíacas, pois, a precisão do método alcançou os 98.65% [20] [21] .

2.5.3 “Support Vector Machine”

A “Support Vector Machine” (SMV) é uma técnica baseada na teoria de aprendizagem estatística, estabelecendo uma série de princípios que devem ser seguidos na obtenção de classificadores com boa capacidade de generalização. A teoria de aprendizagem é baseada em condições matemáticas que auxiliam na escolha de um classificador particular a partir de um conjunto de dados de treino. Essas condições consideram o desempenho do classificador no conjunto de treino e a sua complexidade, com o objetivo de obter um bom desempenho também para novos dados do mesmo domínio. Na aplicação desta teoria, assume-se inicialmente que os dados do domínio em que a aprendizagem ocorre são gerados de forma independente e identicamente distribuída de acordo com uma distribuição de probabilidade que descreve a relação entre os alvos e os seus rótulos.

A técnica SVM surgiu pela aplicação direta dos resultados fornecidos pela teoria da aprendizagem estatística. Esta técnica é maioritariamente utilizada na obtenção de fronteiras lineares ou não lineares para a separação de objetos pertencentes a duas ou mais classes. Geralmente, esta técnica realiza a classificação através da construção de um hiper plano n-dimensional que separa os dados em n categorias de forma otimizada, ou seja, pega nos dados de entrada e classifica-os como uma entre n categorias existentes.

Para que seja eficaz, é necessário utilizar um conjunto de dados de entrada e de saída de treino para construir o modelo de SVM que pode ser utilizado para a classificação de novos dados. A SVM é simplesmente baseada nos valores de entrada ou saída, ou seja, o método é simples, insere uma entrada e, com base no modelo criado através do treino, devolverá uma saída. Por exemplo, na Figura 8, x_1 e x_2 são duas entradas de SVM, o algoritmo mapeia cada conjunto de entradas num espaço com 2 dimensões e dá uma saída de um hiper plano de 1 dimensão que separa os dados de treino em duas classes.

A melhor forma de desenvolver este algoritmo é considerar como sendo um caso bidimensional, pois, assim permite criar uma SVM de duas entradas e retornar uma única saída. A SVM observa os dados num espaço bidimensional e utiliza o algoritmo de regressão para encontrar um hiper plano unidimensional que separa com maior precisão os dados entre as suas duas categorias. À medida que aumenta o número de entradas da SVM, torna-se mais difícil de desenvolver o seu modelo e uma vez que o número de entradas

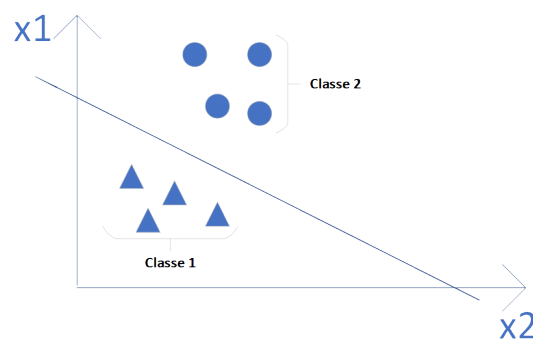


Figura 8: “Support Vector Machine”.

para a SVM varia de um a infinito, a capacidade computacional limita a quantidade de entradas que podem ser utilizadas.

Este algoritmo pode ser usado para desafios de classificação ou regressão, mas é mais utilizado principalmente em problemas de classificação. O benefício de utilizar este método para extrair um padrão complexo dos dados é que não é necessário um conhecimento prévio acerca do comportamento dos dados, pois este é eficaz na análise dos dados e a extrair somente discernimentos e relações. Uma das melhores características deste método é que ele é capaz de resolver os erros e ruídos nos dados de forma atempada. É geralmente eficaz em descobrir qual o padrão de fundo nos dados e filtrar os valores dos dados desnecessários.

O método SVM produz um alto grau de precisão de classificação de doenças cardíacas, uma vez que a média global passa a ser 85% de classificação bem-sucedida [22]. Há também outro estudo, este aplicado à tarefa de resolver problemas de classificação de doenças cardíacas e os recursos extraídos com base em propriedades estatísticas, nesta precisão é sempre alta, pois, o método produziu um alto grau de precisão de classificação, com uma proporção significativamente alta, 93%, de positivos verdadeiros alcançados [23].

2.5.4 Regressão logística múltipla

A regressão logística múltipla é baseada numa técnica estatística e constitui um dos algoritmos de “Machine Learning” mais importantes na análise de dados quando se pretende modelar relações entre variáveis ou entradas. O principal objetivo deste algoritmo é explorar a relação N entradas e N saídas. Este algoritmo tem como objetivo modelar, a partir de um conjunto de dados, a relação “logística” entre N saídas poliatômicas e uma série de entradas (contínuas, discretas).

Este algoritmo é uma poderosa forma estatística de modelar um resultado poliatômico com duas ou mais entradas. Ele mede a relação entre as entradas e as saídas, estimando as probabilidades usando uma função softmax que é a distribuição logística cumulativa. Por exemplo, na Figura 9, x_1 e x_2 apresentam as entradas deste algoritmo, ele organiza cada conjunto de entradas num espaço com 2 dimensões e dá duas

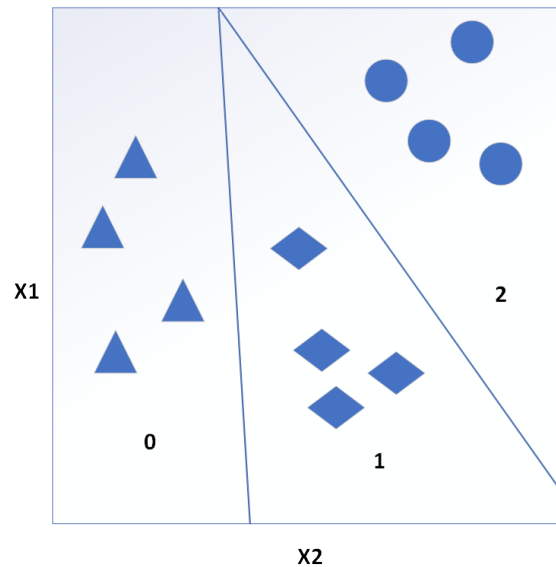


Figura 9: Regressão logística múltipla.

saídas que separam os dados em grupos 0,1 e 3.

De acordo com alguns estudos realizados, verificou-se que o algoritmo Regressão logística múltipla alcançou a precisão de previsão mais alta, sendo esta de 85%, o que significa que o método proposto será útil em diagnósticos de doenças [24].

2.5.5 Random forest

Esta técnica aplica um método da probabilidade para produzir amostras aleatórias de conjuntos de treino para cada árvore. Cada subconjunto novo de treino é construído, com reposição, a partir do conjunto de treino original. A árvore é construída usando este novo subconjunto e uma seleção aleatória de atributos. Para cada nó da árvore, um subconjunto de m atributos é selecionado aleatoriamente e avaliado. O melhor atributo é escolhido para dividir o nó. O valor de m é fixado para todos os nós.

A “Random forest” é uma técnica computacionalmente eficiente que pode operar rapidamente sobre grandes bases de dados. Como o nome sugere, esta técnica cria uma floresta com várias árvores, ou seja, é definida formalmente como um classificador composto por uma coleção de árvores. Cada árvore depende dos valores de uma amostra aleatória e com a mesma distribuição para todas as árvores na floresta. O erro de generalização de uma floresta de classificadores de árvores depende da força das mesmas de forma individual e da correlação entre elas. Normalmente, quanto mais árvores na floresta, mais robusta é a mesma. Da mesma forma que no classificador de florestas aleatórias, quanto maior for o número de árvores na floresta, melhores são os resultados.

Esta técnica tem sido usada em muitos projetos de pesquisa recentes e em aplicações do mundo real

em diversos domínios. Aliás, “Random Forest” tem uma melhor precisão em relação à maioria dos métodos atuais, mostrando ser eficaz para grandes bases de dados. Esta fornece estimativas das variáveis que são importantes na classificação, gerando uma estimativa imparcial interna do erro de generalização [25] [26].

O “Random forest” é aplicado com sucesso na previsão dos fatores de risco de doenças cardíacas, pois é um algoritmo de aprendizado altamente preciso, adequado para aplicações médicas. Conforme alguns estudos, o método apresentou uma precisão de 83,70% para o conjunto de dados de estêreo cardíaco [27].

2.5.6 “Naive Bayes Classifier”

Este algoritmo é baseado nos métodos probabilísticos Bayesianos. Os métodos probabilísticos Bayesianos descrevem a probabilidade de um evento baseado num acontecimento e trabalha com a probabilidade condicional que é a probabilidade de algo acontecer, tendo como base acontecimentos passados. Assumem que a probabilidade de um evento A dado um evento B não depende apenas da relação entre A e B, mas também da probabilidade de observar A independentemente de observar B. A probabilidade de ocorrência do evento A pode ser estimada pela observação da frequência com que este evento ocorre. De forma semelhante, é possível estimar a probabilidade de que um evento B ocorra dado que foi observado o evento A, $P(A|B)$. Nos métodos probabilísticos para problemas de decisão, o objetivo é estimar $P(A|B)$, onde A representa a classe e B o valor observado dos atributos para um exemplo de teste.

Este classificador prevê as probabilidades de associação para cada classe, bem como a probabilidade de que um determinado ponto de dados pertença a uma determinada classe. A classe com maior probabilidade é considerada a classe mais provável. Por exemplo, $P(A)$ representa a probabilidade de ocorrência de uma determinada doença, e $P(B)$ representa a probabilidade de um doente ter um determinado resultado num exame de ecocardiograma. Para um determinado paciente, a variável A não é observável. Perante a evidencia do resultado do exame, pode inferir o valor mais provável de A estimando $P(A|B)$.

O algoritmo “Naive Bayes Classifier” tem uma precisão alta para detetar doenças cardiovasculares, especialmente para adultos. De acordo com o resultado obtido pelo estudo, o algoritmo conseguiu prever doenças cardíacas com uma taxa de precisão de 86,53% [28]. A maioria dos investigadores da altura concordaram que este algoritmo tem a contribuição na área médica, de maneira a apoiar a análise e deteção de doenças cardiovasculares [28].

2.5.7 “K-Nearest Neighbour”

Este é um método baseado em distancias e utilizado com frequência. O conceito deste método é que dados similares tendem a estar concentrados na mesma região do espaço de entrada. Alternativamente, dados que não são similares estarão distantes entre si. O método classifica um novo objeto com base nos

exemplos do conjunto de treino que são próximos desse objeto.

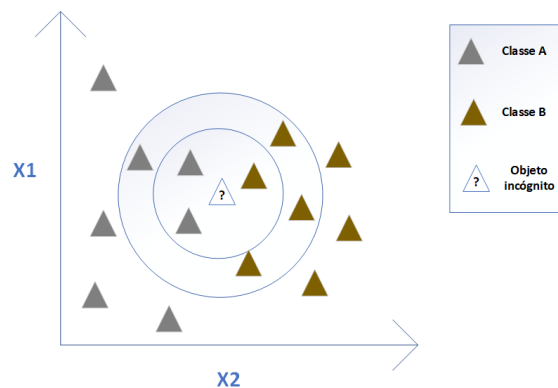


Figura 10: “K-Nearest Neighbour”.

O algoritmo “K-Nearest Neighbour” considera, os k objetos do conjunto de treino mais próximos do ponto de teste X_t , em que k é um parâmetro do algoritmo. Quando o parâmetro K é maior do que 1, para cada ponto de teste, são obtidos K vizinhos. Cada vizinho vota numa classe. No exemplo da Figura 10, para $k=3$, o objeto incógnito seria classificado como pertencendo à classe A, enquanto para $k=5$ o objeto incógnito seria classificado como pertencendo à classe B.

Este algoritmo é o mais simples de todos os algoritmos de aprendizagem automática, não aprende um modelo compacto para os dados, apenas memoriza os objetos de treino.

Os resultados obtidos por um estudo, mostraram que a aplicação de K-Nearest Neighbour num conjunto de dados de doenças cardíacas também é eficaz e alcançou uma precisão, 81,85% [29].

2.5.8 “K-Means”

Este algoritmo é parecido com o algoritmo anterior, sendo usado para fazer agrupamentos automáticos de dados, tendo como base o seu grau de semelhança. O algoritmo é baseado no cálculo da distância de cada elemento em relação aos valores escolhidos pelo utilizador, agrupando o elemento ao grupo que se encontra mais próximo e recalculando o centroide. O processo é repetido até que todos os elementos façam parte de um dos grupos.

O algoritmo “K-means” classifica n objetos dentro K grupos onde cada objeto pertence ao grupo mais próximo do centro do grupo e k é um parâmetro deste algoritmo. A Figura 11 ilustra o funcionamento deste algoritmo. A figura mostra três grupos distintos que foram identificados pelo algoritmo. O algoritmo calcula 3 centroides, cada centroeide corresponde a um grupo e particiona as bolinhas dentro desses três grupos. Como a figura mostra as bolinhas azuis correspondem ao grupo 1, as bolinhas cinzentas correspondem ao grupo 2 e as bolinhas amarelas correspondem ao grupo 3.

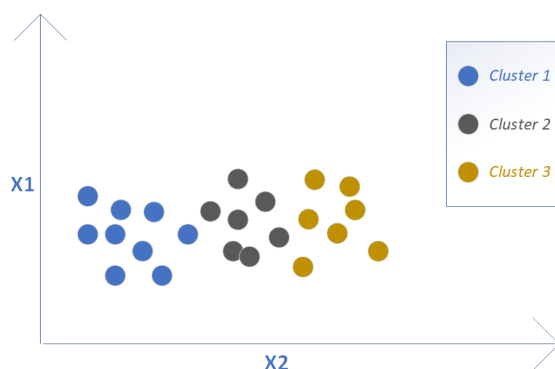


Figura 11: “K_mean”.

Este método é usado principalmente para melhorar a eficiência, sendo mais utilizado para encontrar grupos que não foram classificados explicitamente nos dados. Ele pode ser usado para confirmar os pressupostos de negócios sobre quais tipos de classes existem ou identificar grupos desconhecidos em conjuntos de dados [30]. Uma vez que este método foi executado e os grupos são classificados, qualquer novo dado pode ser facilmente atribuído ao grupo pretendido [31].

De acordo com alguns estudos realizados, o algoritmo “K-means” é um método eficiente de agrupamento para prever pacientes que sofrem de doenças cardíacas, visto que alcançou uma precisão de 80.86%. Este modelo oferece uma solução para qualquer tipo de análise com informações precisas e detalhadas [32].

2.5.9 Conclusão

Com base na revisão acima, pode-se concluir que existe uma enorme variedade para algoritmos de “Machine Learning” que possa ser aplicado na determinação dos tipos de patologias cardíacas. Cada um dos algoritmos mencionados acima teve bons resultados em alguns casos, mas maus resultados noutros casos.

De acordo os estudos realizados, embora todos os métodos de “Machines Learning” tenham tido uma bom resultado, a rede neuronal artificial foi o método que obteve um resultado mais elevada. Foi também concluído em alguns estudos, em que se utilizaram vários métodos de “Machine Learning” entre os quais a rede neuronal artificial[33]. Mas algum estudo mostrou também que os resultados do algoritmo de classificação “Support Vector Machine”, em comparação com a classificação rede neuronal artificial, são muito encorajadores e precisão [23]. O desempenho aprimorado da classificação “Support Vector Machine” deve-se ao fato de que este método pode evitar pesquisas repetitivas para encontrar os dois melhores pontos para usar em cada etapa de otimização. Aplicando ao objetivo de resolver o problema de classificação da doenças cardíacas e as características extraídas com base em propriedades estatísticas, obtêm uma precisão maior no modelo de classificação SVM [23].

2.6 Bibliotecas de “Machine Learning”

Existe uma ampla gama de bibliotecas e “frameworks” de “Machine Learning” disponíveis no mercado e desempenharam um papel fundamental no desenvolvimento de “Machine Learning”. Estas permitem que os utilizadores implementem facilmente o algoritmo de “Machine Learning”, gerem novos projetos e criem novos sistemas de “Machine Learning”. Os frameworks de “Machine Learning” vêm com componentes pré-construídos que são fáceis de entender e codificar. Um bom “framework” reduz a complexidade de definir modelos “Machine Learning”. A seguir, pretenda se mostrar algumas dessas “frameworks”.

2.6.1 “TensorFlow”

Esta é uma biblioteca de computação numérica desenvolvida pela Google para “Machine Learning” aplicável a uma ampla variedade de tarefas. É um sistema para criação e treino de redes neurais para detetar, prever e analisar um evento. Um programa “TensorFlow” está definido em duas partes: construção do grafo computacional e execução do grafo computacional. Um grafo computacional é uma série de operações “TensorFlow” dispostas em um grafo de nós. A execução do grafo, é tipicamente um “loop” de treino.

Os Algoritmos no “TensorFlow” são representados como um grafo de fluxo de dados mantendo um estado. A forma dos dados no “TensorFlow” é representada pela estrutura de tensores. Em termos de álgebra linear, tensores são generalizações de matrizes que consistem num conjunto de valores primitivos configurados numa matriz de N dimensões. Por exemplo, um vetor é um tensor com uma dimensão. Uma matriz é um tensor com duas dimensões. Uma pilha de matrizes é um tensor com três dimensões.

“TensorFlow” pode ser executado em CPU (“Central Processing unit”), GPU (“Graphics Processing unit”) e TPU (“Tensor Processing unit”). O TPU foi desenvolvido pelo Google que é adaptado para o “TensorFlow”, o seu objetivo é ser como um acelerador de inteligência artificial programável para fornecer alta taxa de transferência para o algoritmo, e orientada para a execução do modelo, invés de treino. O programa “TensorFlow” costuma ser significativamente mais rápido em GPU do que em CPU e com a GPU para realizar a aprendizagem é mais eficaz. Além disso, o programa está agora disponível para Linux, MacOS, Windows e plataformas de computação moveis, incluindo android e iOS [34].

“TensorFlow” suporta uma computação distribuída, permitindo que partes do gráfico sejam computadas em diferentes processos, que podem estar em servidores completamente diferentes. Além disso, também pode ser usado para distribuir a computação para servidores com GPUs poderosas e ter outro algoritmo feito em servidores com mais memória.

Algumas das vantagens [35] que mostram o “TensorFlow” se tornar mais saliente em relação às outras bibliotecas de Machine Learning são :

- Suporte utilização de GPUs(Unidade de processamento gráfico) para maior paralelização dos processamentos;
- Permite treinar redes neurais colossais com bilhões de dados;
- Fornece API (“Application programming interface”) de alto nível, que tornam a implementação de treinamento de “Machine Learning” extremamente simples;
- Suporta várias funções de otimização com implementação de diferenciação automática. Assim, facilita ao utilizador;
- Suporte à computação distribuída;
- Tem uma ótima ferramenta de visualização chamada “TensorBoard”.

2.6.2 “Apache Mahout”

“Apache Mahout” é uma biblioteca que permite implementações de algoritmos de “Machine Learning” cujos principais objetivos são: processar recomendações, classificações e agrupamentos. Mantido pela “Apache Software Foundation”, o “Apache Mahout” nasceu em 2008 como um subprojecto do “Apache Lucene”, uma outra ferramenta de código destinada a problemas de recuperação de informações.

Esta técnica é muito explorada pelo “data mining” e aplica principalmente métodos estatísticos e probabilísticos, tal como “Naive Bayes Classifier”. Embora não seja uma técnica nova, esta encontra-se em pleno crescimento. Grandes corporações como a Amazon e o Facebook, utilizam algoritmos deste tipo em muitas das suas aplicações tais como: jogos, sistemas de deteção, análise de dados, entre outros. Estes algoritmos são também muito comuns em sistemas de recomendação, como os da Amazon e Netflix, que sugerem novos produtos aos utilizadores com base em compras anteriores. Os algoritmos suportados pelo “Apache Mahout” dividem-se em quatro grupos: filtros colaborativos, classificadores, agrupadores e redutores de dimensionalidade. Todos eles podem ser utilizados através da API [36]. “Apache Mahout” oferece ao utilizador uma estrutura preparada para realizar implementações de “Machine Learning” em grandes volumes de dados. Fornece aplicações para analisar grandes dados de forma eficaz e em tempo real. Para além disso, suporta diversos algoritmos tais como “K-Mean”, “Baye Naive”, entre outros [37].

2.6.3 “Scikit-learn”

A “Scikit-learn” é uma biblioteca de “Machine learning” que utiliza linguagem de programação “Python”. Esta biblioteca permite utilizar vários métodos de classificação, regressão e de agrupamento, como por exemplo, os “Support Vector Machine”, “Random Forest” e “K-means” e possui vários métodos de pré-processamento, como redução de dimensionalidade, transformações de texto e muito mais. É projetada para interagir com as bibliotecas “Python” numéricas. Relativamente à sua origem, esta biblioteca começou com o nome “scikits.learn”, um projeto de Google Summer of Code fundado pelo “David Cournapeau”.

O “Scikit-learn” foi adotado por uma infinidade de marcas bem sucedidas tais como “Spotify”, “Evernote” e “Booking.com”. É uma biblioteca robusta e foca-se essencialmente na facilidade de uso, permitindo que “Machine Learning” realize uma aprendizagem mais rápida. Mas a sua aplicação é mais adequada para a produção, caracterizada por tempo limitado e pelo recurso humano [38].

2.6.4 “Accord.NET”

A “Accord.NET” é uma framework de “Machine Learning” para computação científica, bibliotecas múltiplas para diversas aplicações tais como o processamento de dados estatísticos, reconhecimento de padrões, álgebra linear, redes neurais artificiais, processamento de imagem e sinal, etc [39]. Esta framework é combinado com bibliotecas de processamento de áudio e imagens. Normalmente é utilizada para processamento de sinal, aplicações de visão e tratamento das imagens.

É um “framework” bem documentado que facilita o processamento de áudio e imagem. A ferramenta pode ser usada para otimização numérica, redes neurais artificiais e visualização.

2.7 Conclusão do estado da arte

Neste capítulo foi possível verificar que doenças cardiovasculares são principal causa de morte no mundo, mas existem diversos métodos diagnósticos que podem ajudar a pessoa para prever estas doenças. Relativamente aos métodos enunciados, o método ecocardiograma apresenta os resultados mais preciso e a sua possibilidade da análise do sistema cardíaco em tempo real pelo médico fazem com que este exame tenha bem mais eficiência do que outros mas apresenta alto custo. O Método sismocardiograma oferece bastante informação sobre comportamento cardíaco em vários aspetos. Além disso, o conjunto de parâmetros de um sinal de sismocardiograma, tais como: a forma de onda, tempo de propagação do sinal e a amplitude, tendem a variar consoante com o movimento do coração. Isso torna o método bastante oportuna para a avaliação do comportamento cardíaco e prevenção de doenças cardíacas.

Ao longo desta última década, tem surgido vários algoritmos de “Machine Learning” na previsão de

patologia cardíaca. O algoritmo de rede neural artificial tem melhor desempenho que os outros algoritmos, devido à sua funcionalidade ser baseada na estrutura do sistema nervoso humano. O algoritmo de rede neural artificial apresenta grande poder de representação de conhecimento, devido à criação de relações ponderadas entre as entradas do sistema. Mas o algoritmo de “Support Vector Machine” apresenta também alta precisão da classificação de doenças cardíacas e em algum estudo mostrou que o desempenho da classificação “Support Vector Machine” é melhor do que o algoritmo de rede neural artificial.

Em relação as bibliotecas de “Machine Learning” mostrados anterior, “TensorFlow” apresenta mais vantajosa, a partir desta biblioteca é possível construir os modelos de “Machine Learning” com facilidade e rapidez. Além disso, esta biblioteca oferece várias funções de otimização com implementação de algoritmos. Isto torna a investigação da “TensorFlow” e a implementação de algoritmos bastante oportuna para a análise dos dados cardíacos e a determinação de doenças cardíacas.

3 Análise

Este capítulo analisa o sistema geral, nomeadamente a sua definição e identificação das principais métodos e tarefas ao seu desenvolvimento. Tendo em conta que o componente central deste sistema é implementação de um algoritmo que tenha capacidade para analisar e prever a patologia cardíaca baseada nas técnicas de “Machine Learning” e dados da SCG, torna-se essencial analisar o seu princípio de operação. Para além disso, é necessário definir os métodos fundamentais que identifiquem sinais da SCG, e que garantem a extração das características a partir de dados da SCG para calcular os parâmetros necessários. Para que seja realizado um processamento dos dados a partir das medições provenientes do dispositivo SCG, torna-se necessário definir também os requisitos do sistema da aquisição. Ao fim disso, pretende-se definir um método para determinar alguma correlação entre dados da SCG e do ECO, como já foi referido anterior, combinação entre as ondas da SCG pode oferecer outro tipo de informação útil e esta informação pode corresponder a algum parâmetros do ECO.

3.1 Arquitetura do Sistema

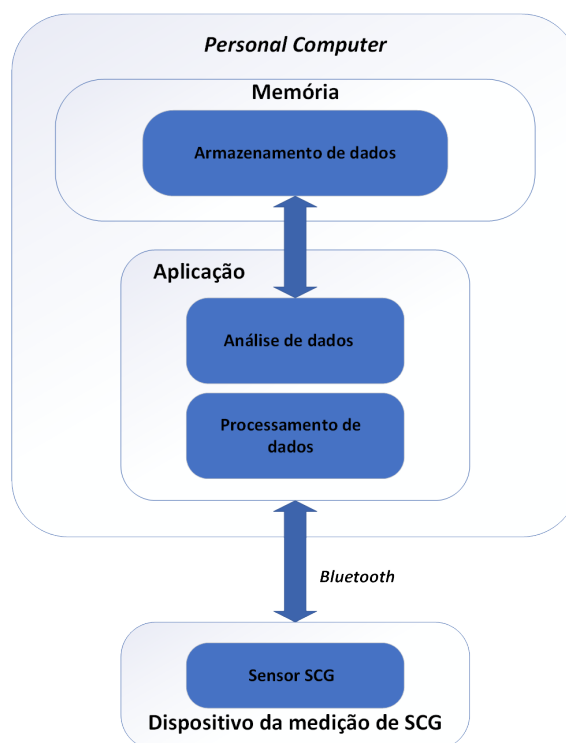


Figura 12: Funcionalidades gerais do sistema.

Na Figura 12 é apresentada uma visão geral do sistema para melhor percepção do conteúdo do trabalho. A arquitetura está dividida em três componentes: o dispositivo da medição da SCG, a aplicação e a memória,

sendo a aplicação é o princípio do componente.

De acordo com o sistema apresentado, pretende-se que o dispositivo da medição estabeleça uma comunicação com o sistema de aplicação realizando a leitura, análise e armazenamento dos dados.

Para que seja possível a observação e análise instantânea do sinal da SCG, torna-se necessário o processo de envios dos dados medidos para aplicação através de “Bluetooth”. Depois, os dados vão ser processados e visualizados no ambiente da aplicação. A aplicação vai ser executada num “Personal Computer” onde esta vai ser implementada. No final, o armazenamento de dados do “Personal Computer” torna-se um requisito para que posteriormente seja possível a realização de uma análise mais cuidada e em maior detalhe dos dados.

3.2 Funcionalidades da aplicação

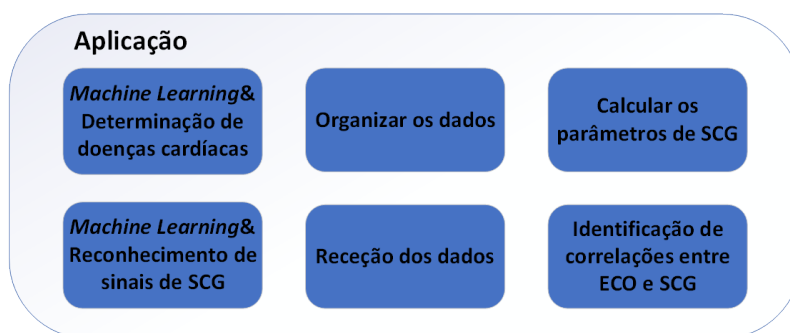


Figura 13: Funcionalidades da aplicação.

Como já foi mencionado anteriormente, a aplicação é o principal componente deste trabalho, por isso, pretende-se mostrar mais em detalhe. A Figura 13 representa as funcionalidades da aplicação que vão ser desenvolvidas durante a dissertação e a relação entre as funcionalidades é dependente uma da outra. Esta aplicação é constituída por seis funcionalidades fundamentais e estão organizadas em seguinte ordem:

- A receção de dados proveniente do dispositivo da medição da SCG.
- O reconhecimento de sinais da SCG, confirmando se o sinal recebido é válido.
- O cálculo dos parâmetros da SCG, obtendo as características a partir de sinais confirmados como batimento cardíaco e taxa respiratória.
- A determinação de correlação entre os dados da SCG e do ECO, descobrindo padrões associados aos dados da SCG e do ECO.
- A determinação de doenças cardíacas, avaliando o estado do coração do paciente.
- A organização de dados, criando ficheiros de armazenamento para cada paciente.

3.3 Comportamento do Sistema

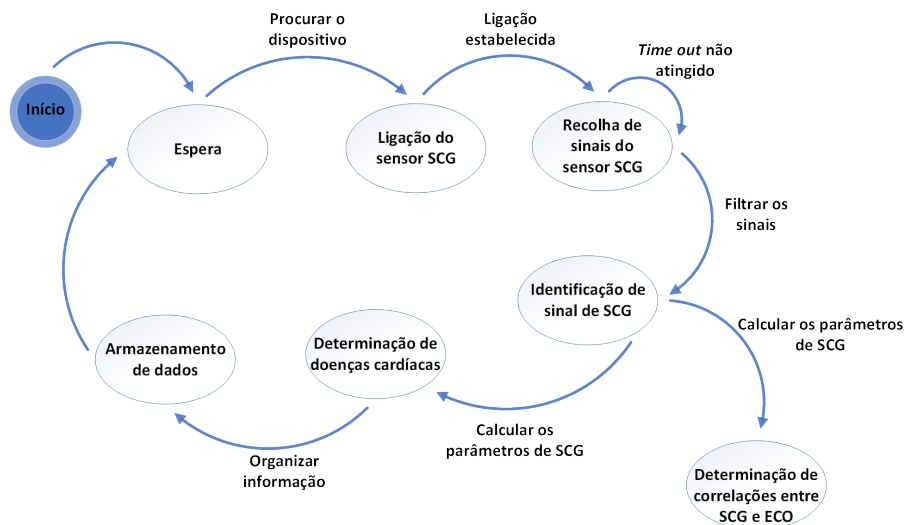


Figura 14: Diagrama do estado do sistema.

A Figura 14 representa o diagrama de estado do sistema de medição e análise. Após a inicialização do sistema, o sistema fica em espera procurando o dispositivo do sensor SCG. Quando é encontrado o dispositivo, este será conectado, sendo de seguida adquirido o sinal da SCG durante tempo definido. Uma vez que o sinal pode conter algum componente indesejado, torna-se necessário processar ou filtrar o sinal recebido. A seguir, o sistema identifica o sinal da SCG a partir do sinal filtrado. Depois, calcula os parâmetros a partir do sinal da SCG e estes parâmetros obtidos vão ser utilizados na determinação de doenças cardíacas e identificação de alguma correlação entre SCG e ECO. Por fim, organiza o resultado obtido e armazena-o. Após tudo isto, o sistema volta a ficar em espera.

3.4 Aquisição de Sinais de SCG

De forma a observar e avaliar os sinais provenientes do sensor SCG, torna-se necessário definir um método de análise que permite obter os parâmetros necessários a partir de sinais da SCG para determinar eventos no comportamento cardíaco. Depois, estes parâmetros serão utilizados na determinação de doenças cardíacas e alguma correlação entre ECO e SCG.

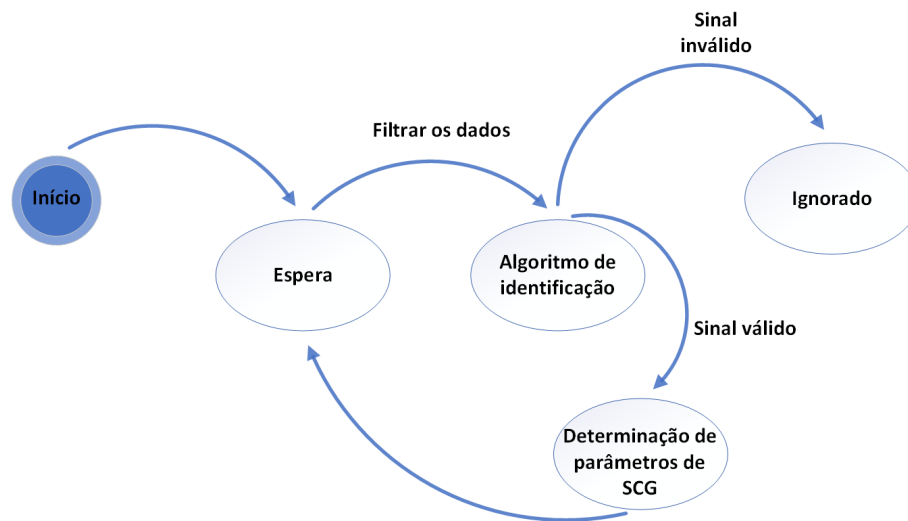


Figura 15: Diagrama do estado da eliminação das sinais inválidos.

A Figura 15 representa o diagrama de estado da eliminação das sinais inválidos da SCG. Após a inicialização deste subsistema, o subsistema filtrando os sinais recebidos. A seguir, executa o algoritmo da identificação sobre estes sinais. Esse algoritmo identifica sinais como válidos e inválidos. Os sinais inválidos vão ser ignorados ou eliminados e os sinais válidos vão ser utilizados na determinação de doenças cardíacas e alguma correlação entre ECO e SCG. Antes de serem utilizados estes sinais, é necessário calcular os seus parâmetros, como batimento cardíaco, uma vez que estes são os fatores mais importante ou fundamentais para determinar as doenças cardíacas. Após tudo isto, o subsistema volta a ficar em espera.

3.4.1 Filtro de Sinais proveniente do Sensor SCG

O desempenho de vários sinais de interferência que têm um grande impacto no sinal da SCG. Por exemplo, a interferência causada pela respiração, quando o corpo humano respira, certos movimentos ocorrerão nos órgãos e tecidos do tórax, e também um componente DC associado à aceleração da gravidade. Isso tudo afetará o sinal da SCG e consequentemente a funcionalidade de reconhecimento de sinais da SCG.

Por isso, é necessário filtrar os sinais para eliminar os componentes de baixa frequência provenientes da respiração, ou ainda os componentes de maior frequência associados à movimentação involuntária do indivíduo e o componente DC associado à aceleração da gravidade.

O filtro é um método de processamento de sinais que tem sido aplicado com bastante sucesso de cancelamento do ruído ou sinal de interferência. A ideia básica do filtro adaptativo é receber um vetor de entrada, estimar um erro e obter uma resposta desejada, que é usado para controlar um conjunto de coeficiente de um filtro. Neste momento, existem duas técnicas de filtros aplicados no sinal da SCG com muito sucesso: Filtro passa-banda e Transformada Discreta de “Wavelet” [40] [41].

3.4.1.1 Passa-banda

O filtro passa-banda aplica uma operação matemática linear a um sinal de entrada. É um método muito utilizado para eliminar as frequências não desejadas de um sinal de entrada ou para selecionar uma frequência de um sinal.

Desta forma, aplica o método a sinal proveniente do sensor SCG para permitir a passagem das frequências de uma certa faixa e rejeita as frequências fora dessa faixa. As frequências estão fora dessa faixa correspondem aos ruídos e outras são sinais da SCG.

3.4.1.2 Transformada Discreta de Wavelet

A “Wavelet” é um tipo da onda de curta duração com um valor médio nulo e existem inúmeras famílias de “Wavelets” e as famílias mais conhecidas e utilizadas são: “daubechies”, “haar” e “symlets”. Desta forma, no contexto desta dissertação, “symlet” será utilizada. A “Symlet” é uma “Wavelet” quase simétrica e é como uma modificação à família db. A motivação de escolher essa é porque “Symlet” tem grande capacidade para evitar distorções desnecessárias ao processar o sinal.

A Transformada Discreta de “Wavelet” é um avanço recente em processamento de sinais. Este método tem vantagem sobre análises espectrais clássicas, visto que permite analisar a periodicidade de sinais em diferentes escalas da variabilidade temporal e não necessita de uma série estacionária. Assim, o método é apropriado para analisar sinais irregularmente distribuídos.

O método é baseado no processo de decomposição. O processo pode ser iterado, com sucessivas aproximações a serem decompostas, de forma a que um sinal seja dividido em várias componentes. A decomposição pode ser efetuada até que os detalhes individuais consistam numa amostra simples, assim, facilita a análise de sinais e remove alguns componentes que não são considerados, ou seja ruído.

Depois da decomposição, é necessário reconstruir o sinal sem perder a informação face ao sinal original. Para sintetizar um sinal, este é reconstruído a partir dos coeficientes de “Wavelete”.

Deste modo, o método proposto também pode ser usado para processar o sinal da SCG. Decompondo o sinal da SCG em vários sinais pequenos e em diferentes escalas, sendo analisados os sinais pequenos e removidos os sinais de ruído, por fim, são calculados os coeficientes de “Wavelete” e reconstruído o sinal.

3.4.2 Identificação de Sinais da SCG

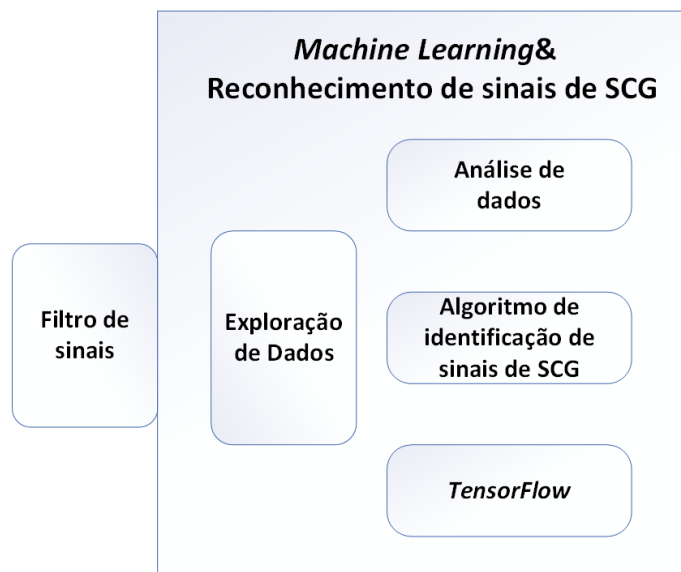


Figura 16: Identificação de sinais da SCG.

Na Figura 16 está apresentada uma esquemática de “Machine Learning” para identificar sinais da SCG. Uma vez que o sinal pode conter algum componente inválido que não corresponde o movimento do coração, isso pode ser devido ao movimento ou comportamento indesejado proveniente do humano ou má medição pelo sensor SCG e este componente não é possível ser eliminado por filtro. Como já foi mencionado no estado da arte, o sensor SCG convencionou um método que tira proveito dos movimentos produzindo um sinal do corpo, ou seja, o sensor é muito sensível ao movimento.

De acordo com a esquemática apresentada, pretende-se que o sinal proveniente do sensor seja filtrado antes de ser utilizado, como já foi explicado anteriormente, o sinal pode conter algum ruído e este ruído pode afetar a funcionalidade de “Machine Learning”.

Para que facilite o trabalho de identificação, é necessário utilizar um método da exploração de dados para obter padrões e tendência na análise exploratória. A seguir, a técnica de “Machine Learning” analisa essa informação e utiliza o algoritmo da identificação para identificar os sinais da SCG, neste caso o algoritmo vai ser implementado com ajuda da biblioteca “Tensorflow”. Concluindo, os dados identificados no seu formato original vão ser guardados no armazenamento para que posteriormente seja possível a realização de uma análise mais cuidada e em maior detalhe ao sinal adquirido.

3.4.3 Exploração de dados

A informação obtida na análise exploratória de dados pode ajudar na construção de modelos de “Machine Learning”. Uma das formas mais simples de explorar o conjunto de dados da SCG consiste em extrair

estatísticas descritivas.

Estas assumem que os dados são gerados por um processo estatístico. Como o processo caracteriza-se por vários parâmetros, as medidas podem ser interpretadas como estimativas dos parâmetros estatísticos da distribuição normal que gerou as informações.

Desta forma, no contexto desta dissertação, pretende-se utilizar estas medidas para capturar informação como centralidade e dispersão.

A medida de dispersão mede a variabilidade de um conjunto de dados. Permite averiguar se os dados se encontram amplamente dispersos ou relativamente concentrados em torno de um valor, que pode ser SCG ou não. A medida de dispersão que vai ser utilizada nesta dissertação é o desvio padrão.

A medida da centralidade mede pontos de referência nos dados. Para esta dissertação, a medida da centralidade vai ser a média, uma vez que esta é um bom indicador do valor central de um conjunto de valores.

Antes de aplicar o método da exploração de dados, é necessário partir o sinal recebido em vários sinais pequenos, visto que algum segmento do sinal não é válido e é necessário removê-lo.

3.4.4 Algoritmos de identificação das sinais da SCG

Para que seja possível identificar sinais da SCG e remover sinais inválidos, é também utilizando algoritmos de “Machine Learning”. A “Machine Learning” além de ter capacidade de ajudar prevenir os tipos de doenças cardíacas, também é capaz de ser aplicados com sucesso de identificação dos sinais.

Os algoritmos de “K-Mean” e regressão logística múltipla são eficazes em resolver o problema da classificação ou identificação de dados. Ambos são fáceis de entender e produzem resultado com alta precisão. Com estas vantagens, pretende-se implementar estes dois algoritmos e comparar os resultados obtidos para verificar qual método é mais adequado na identificação dos sinais da SCG.

3.4.4.1 “K-Mean clustering”

“K-Mean” é um dos muitos métodos de “Machine Learning” não supervisionado. Como já foi mencionado no estado da arte, este método é baseado na distância de cada elemento em relação aos valores de centroide e agrupa o elemento que se encontra mais próximo. Os elementos presentes no trabalho são o desvio padrão e a média do sinal da SCG. Cada grupo formado representa uma classe. Neste caso, as classes correspondem a um sinal válido e sinal inválido.

O método vai analisar todos os dados da SCG e criar classificações. Isto é, o algoritmo vai indicar uma classe e vai prever quais dados pertencem a esta classe.

$$distancia = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

1: Equação do cálculo da distância

$$centroide = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

2: Equação do cálculo da centroide

As Equações 1 e 2 apresentadas são as duas equações necessárias para este método. A Equação 1 é para calcular a distância entre os dados num espaço. A Equação 2 é para descobrir o centro de cada área (cada área corresponde uma classe).

Para gerar as classes e classificar os dados da SCG, o método gera k centroides em localizações aleatórias e calcula as distâncias desses centroides em relação aos dados da SCG. Geralmente utiliza-se Equação 1 para calcular a distância entre cada dado e centroide num espaço de duas dimensões. Após o cálculo da distância utiliza a Equação 2 para calcular centroides para cada uma das classes e classifica os dados de acordo com a sua distância dos centroides. Conforme o algoritmo vai iterando, o valor de cada centroide é ajustado pela média dos valores de cada dado que pertence a este centroide.

Para simplificar a explicação de como o método vai ser aplicado no trabalho, este vai ser apresentado em cinco passos:

- Gerar valores aleatórios para os centroides, uma vez que é necessário atribuir os valores iniciais aos k centroides.
- Calcular a distância entre cada dado e os centroides das classes.
- Classifica cada dado nas classes de acordo com a sua distância do centroide de classe
- Calcular os novos centroides para cada classe.
- Voltar para o passo 2 repetindo iteramente o ajustamento do valor de cada centroide pela média dos valores de distância entre dados e classes.

3.4.4.2 Regressão logística múltipla

A regressão logística múltipla é um método de “Machine Learning” supervisionado e trata-se de um modelo de regressão para variáveis dependentes ou de resposta multinomialmente distribuídas. Nesta

dissertação, o método funciona de forma muito semelhante à regressão linear, mas com uma variável de resposta trinomial. Os atributos presentes neste método são o desvio padrão e a média do sinal da SCG e cada distribuição representa uma classe. Neste caso, as classes correspondem a frequência normal, frequência alta e frequência baixa. A classe frequência normal classifica todo o conjunto de dados que correspondem ao ritmo do coração, ou seja, sinal da SCG, e estas duas últimas classes identificam todo o conjunto de dados que não corresponde ao ritmo do coração, ou seja, é considerado um sinal inválido.

O modelo deste método pode ser construído pelo sistema, que, através de séries de treino, permite criar 3 regiões de fronteira com cada uma correspondendo a uma classe.

3.4.4.2.1 Função “Softmax” A função “Softmax” é uma generalização da função logística, sendo um elemento principal deste algoritmo. Esta é usada para representar distribuições categóricas, isto é, uma distribuição de probabilidade sobre N resultados possíveis diferentes.

De maneira que seja possível realizar a identificação de dados da SCG, dada uma entrada de sinal, deseja-se que a hipótese calcule a probabilidade de que $P(y = k | \text{sinal})$ para cada valor de $k = 1, \dots, K$. Ou seja, pretende-se estimar a probabilidade do rótulo de classe assumir cada um dos diferentes K valores possíveis. Assim, a hipótese produzirá um vetor K-dimensional (cujos elementos somam 1) dando as probabilidades estimadas de K. Mais concretamente, a hipótese $h(\text{sinal})$ assume a forma:

$$\text{Softmax}(x_j) = P(y = j|x) = \frac{\exp(x_j)}{\sum_{j=1}^K \exp(x_j)} \quad (3)$$

3: Equação de “Softmax”

3.4.4.2.2 Função “Cross Entropy” A “Cross Entropy” é uma função usada para quantificar a diferença entre duas distribuições de probabilidade sobre o mesmo conjunto de eventos. Neste caso, a função vai ser usada para definir uma função “loss” em “Machine Learning” e otimização. A real probabilidade é o rótulo verdadeiro, e a distribuição dada é o valor previsto do modelo. Além disso, a função lida com a classificação de um conjunto de dados em três classes possíveis genericamente rotuladas 0, 1 e 2. O modelo prevê, assim, uma saída 0, 1 ou 2, dado um vetor de entrada.

$$\text{Cross_Entropy}(y_i, x_i) = - \sum_{i=1}^n y_i \log(\text{Softmax}(x_i)) \quad (4)$$

4: Equação de “Cross_Entropy”

A Equação 4 ilustra a função “Cross Entropy” para as distribuições x_i e y_i sobre um determinado conjunto.

3.4.4.2.3 Função “loss” A Equação 5 representa a função “loss”. A Função “loss” é um método de avaliação de como o algoritmo modela o conjunto de dados. Se a sua previsão for pior, a função produzirá um número maior, caso contrário, resultará num número menor.

$$Loss = Mean(y_i \log(Softmax(x_i))) \quad (5)$$

5: Equação de “loss”

3.5 Cálculo dos parâmetros da SCG

Na Figura 17 é ilustrado um esquema da funcionalidade do cálculo dos parâmetros da SCG. Depois de identificar os sinais da SCG, é necessário extrair as informações ou as características importantes a partir dos sinais. Por exemplo, o batimento cardíaco, taxa respiratória, as ondas típicas da SCG e outros. Como pode observado esta figura, a função faz leitura de sinais da SCG, determina os parâmetros necessários e armazena-os para que possam ser utilizados no diagnóstico de doenças cardíacas. Para melhor percepção deste conteúdo, a Figura 18 apresenta uma visão geral do processo do cálculo.

De acordo com a Figura 18, para obter os parâmetros a partir do sinal da SCG, é necessário primeiramente determinar as ondas AO para obter a onda média.

De seguida, pretende-se a implementação de métricas para análise desta onda média que permitam obter os parâmetros sobre o comportamento cardíaco e investigar a aplicabilidade deste sinal da SCG para diagnóstico de patologias cardíacas. Adicionalmente, o método do filtro passa-banda permite distinguir a frequência respiratória, enquanto que a determinação das ondas AO permite obter a frequência cardíaca.

Os parâmetros dos intervalos de tempo podem ser obtidos por uma combinação entre duas ondas distintas. Como já foi mencionado no estado da arte, a combinação entre dois picos das duas ondas pode fornecer outro tipo de informação útil para analisar o comportamento do coração como o intervalo de tempo entre AO e AC podendo corresponder um tempo de ejeção sistólica. Por isso, durante desenvolvimento pretende-se explorar mais informações sobre o comportamento do coração que o sinal da SCG possa oferecer.

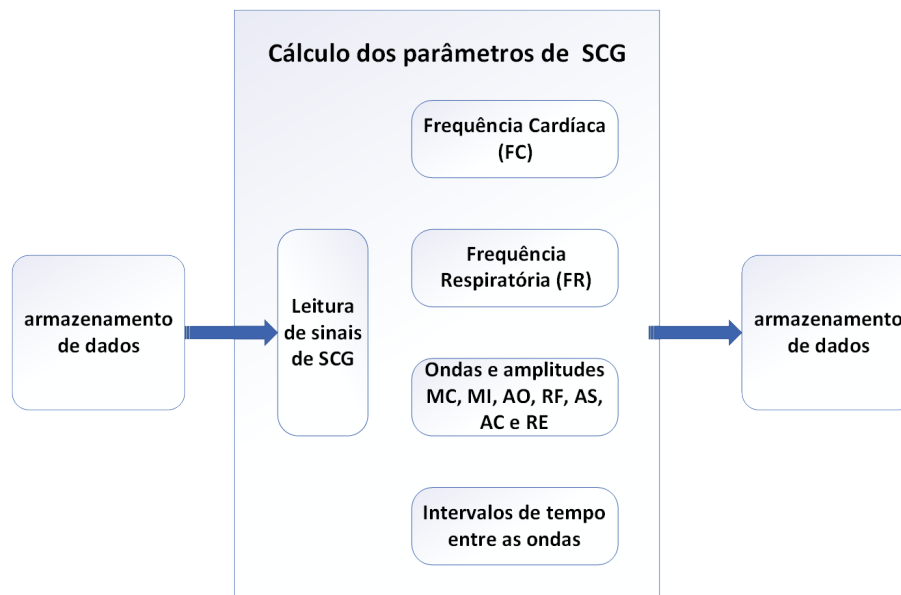


Figura 17: Sistema do cálculo dos parâmetros da SCG.

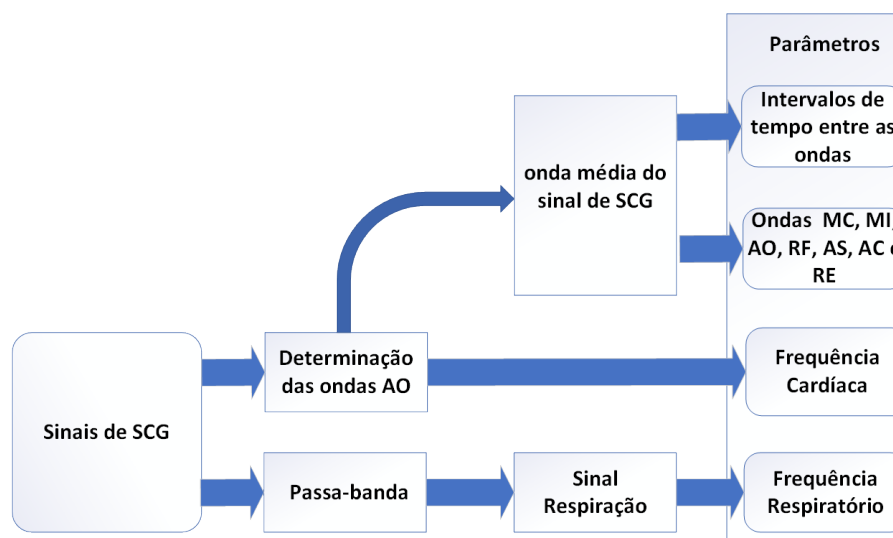


Figura 18: Método do cálculo dos parâmetros da SCG.

3.5.1 Detecção das ondas AO

A detecção das ondas típicas da SCG, como MC, MI, AO, RF, AS, AC e RE, é o ponto de partida para obter características da SCG. Como já foi mencionado no estado da arte, as ondas representam a atividade aquando da contração dos ventrículos, sendo que, por si só, já transmite muita informação acerca do estado do coração e, consequentemente, na identificação de anormalidades do coração.

A onda AO é a que mais se salienta e mais fácil de ser detetada, pois apresenta um amplitude maior do que outros. Com a deteção desta onda é possível identificar algumas doenças, tal como ritmo cardíaco anor-

mal. Além disso, a realização desta detecção é possível indicar algum parâmetro importante como batimento cardíaco e as outras ondas típicas da SCG. Depois do descobrimento da onda AO, é necessário realizar a sobreposição dos vários intervalos identificados por esta e calcular o valor médio do conjunto de cada amostra sobreposta para determinar um sinal que corresponde média das curvas medidas dos sinais da SCG, assim permitindo uma análise mais concreta ao conjunto de ondas e a identificação das características dos sinais da SCG.

Nas últimas décadas o avanço tecnológico tem proporcionado o desenvolvimento de novos métodos que permitem a execução de algoritmos de detecção da onda AO da SCG cada vez mais eficazes. Existem diversos métodos utilizados na detecção da onda AO, nomeadamente método baseado em envelope [42] [43].

3.5.1.1 Envelope

O envelope é um método de modulação que mede a variação de uma das características de um sinal ao longo do tempo, como a amplitude e a frequência. Dado uma onda modulada, o método utiliza esta onda para realizar a modulação do sinal e o seu resultado é uma onda da frequência igual ao portador, em que a amplitude varia ao longo do tempo, com a mesma frequência e forma de onda modulada.

Para que seja possível ser aplicado o método no ambiente do trabalho, é necessário utilizar o algoritmo “Hilbert transform” para obter o método. Este algoritmo é uma operação linear que gera uma função na saída no mesmo domínio e recebe uma função ortogonal na entrada. Neste caso, o algoritmo funciona como um filtro linear sobre as componentes de frequência do sinal original, mais especialmente na sua fase. A fase dos componentes negativos é deslocada de $+90^\circ$ e a fase dos componentes positivos mantêm-se. As amplitudes dos componentes não se alteraram.

Desta forma, no contexto deste trabalho, a partir do “Hilbert transform” é determinado o envelope do sinal da SCG que servirá para detetar a onda AO. A determinação desta onda realiza-se com base na determinação valores superiores a um “threshold” baseando no valor médio do envelope. Assim, através da identificação das ondas AO, torna-se possível determinar algum parâmetro cardíaco e outro tipo de ondas que está associado com ela.

3.6 Correlação entre os dados do ECO e da SCG

De acordo com a Figura 19, pretende-se que desenvolva uma função para determinar alguma relação possível entre os dados do ECO e da SCG. Como já foi mencionado no estado da arte, as combinações entre as ondas do sinal da SCG podem oferecer outros tipos das informações, e estas informações são possíveis identificadas através de correlação com os dados do ECO. Por isso, é importante para avaliar a relação entre ECO e SCG, assim, permite descobrir as características da sua consistência.

Esta função realiza a leitura dos dados do ECO e da SCG, processa os dados e identifica uma correlação possível entre ECO e SCG.

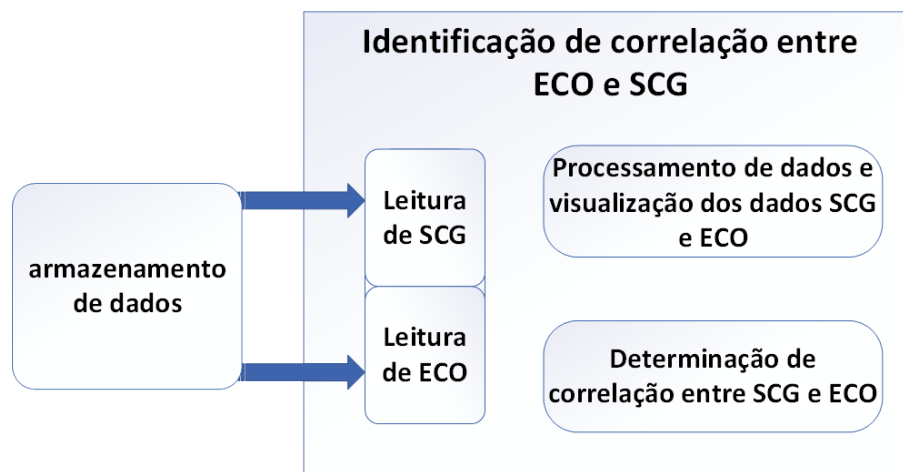


Figura 19: Sistema para identificar uma correlação entre ECO e SCG.

Para que seja possível uma observação dos dados do ECO e da SCG, torna-se necessário o processamento dos dados recebidos e visualização das informações. De forma a determinar a relação entre os dados do ECO e da SCG, tenciona-se a implementação de maneiras para análise dos dados que permitam identificar uma correlação entre ECO e SCG.

3.6.1 Método de correlação de “Pearson”

O algoritmo “Pearson” é um método da correlação estatística para avaliar um possível grau da correlação linear entre duas variáveis de escala métrica [44]. O seu valor assume apenas números entre -1 e 1, que reflete a intensidade de uma relação linear entre dois conjuntos de eventos [45]. De seguida, pretende-se mostrar mais em detalhe.

- 0.9 a 1 positivo ou negativo indica uma correlação muito forte;
- 0.7 a 0.9 positivo ou negativo indica uma correlação forte;
- 0.5 a 0.7 positivo ou negativo indicia uma correlação moderada;
- 0.3 a 0.5 positivo ou negativo indica uma correlação fraca;
- 0 a 0.3 positivo ou negativo indica uma correlação desprezível.

Desta forma, no contexto deste trabalho, o algoritmo destina-se a descobrir uma relação possível entre os dados do ECO e da SCG.

3.7 Método de classificação de doenças cardíacas

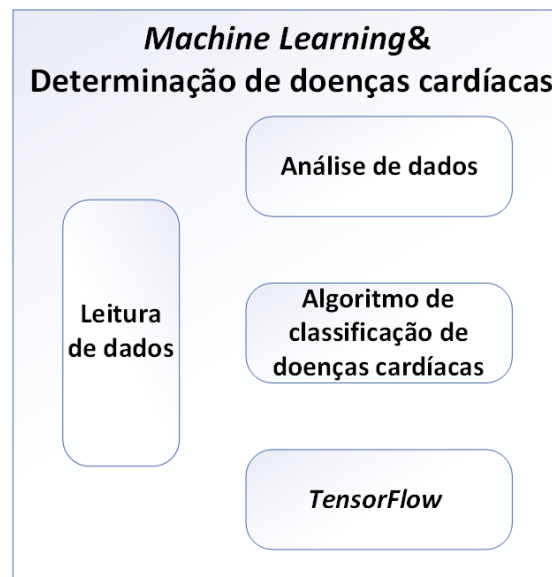


Figura 20: Sistema da determinação de doenças cardíacas.

Na Figura 20 está apresentada uma estrutura da identificação automática das doenças cardíacas baseada nas técnicas de "Machine Learning".

Conforme o diagrama apresentado, pretende-se que o sistema de "Machine Learning" tenha capacidade de analisar os dados e aplique um algoritmo da classificação dos dados para fazer prevenção sobre a possível ocorrência dos problemas cardíacos. A biblioteca "TensorFlow" apresentada no diagrama vai ser uma ferramenta para desenvolver este algoritmo com facilidade. Por fim, os resultados obtidos serão guardados no armazenamento.

3.7.1 Diagrama de estados

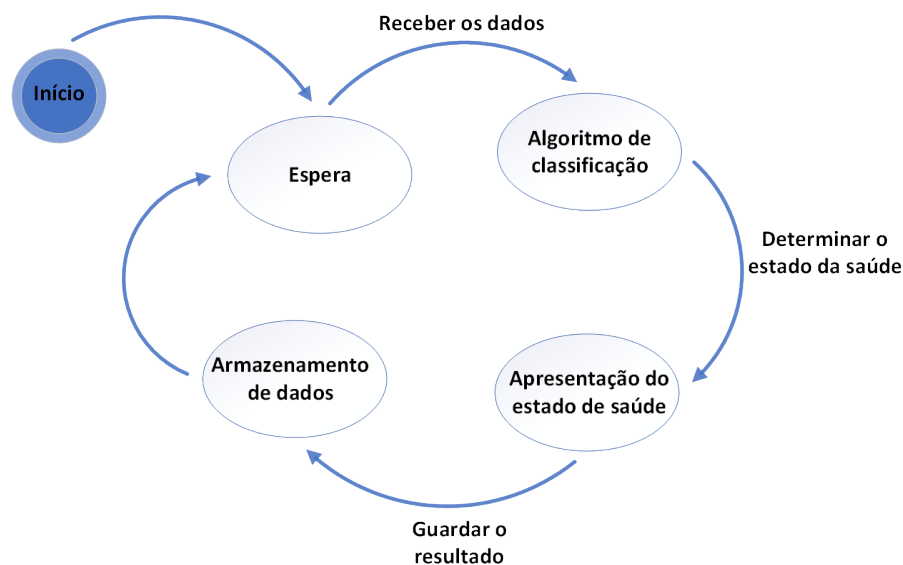


Figura 21: Diagrama de estados.

A Figura 21 apresenta o diagrama da determinação do estado do coração. Ao iniciar o subsistema, este fica em espera, lendo dados a partir do armazenamento dos dados. Os dados serão processados, sendo de seguida executado o algoritmo da classificação sobre os mesmos. Se o algoritmo detetar alguma anomalia cardíaca, será informado de uma possível doença cardíaca. Caso não seja detetada nenhuma doença cardíaca, será informado que a coração do paciente está saudável. Por fim, todos os resultados concluídos serão guardados no armazenamento de dados.

3.7.2 Comparação de ECO e SCG

A análise das características presentes num conjunto de dados permite a descoberta de padrões e tendência que podem fornecer informações valiosas que ajudem a precisão da identificação de doenças cardíacas. Muitas dessas características dependem do tipo de dados escolhidos. Como já foi indicado no estado da arte, os dados de ECO apresentam uma precisão mais alta em relação aos outros, mas os dados da SCG oferecem mais características sobre o comportamento do coração. Logo, pretende utilizar estes dois tipos de dados para concluir qual será tipo de dados ideal para “Machine Learning” realizar uma aprendizagem.

3.7.3 Algoritmos de prevenção de doenças cardíacas

Como já foi mencionado no estado da arte, existem vários algoritmos disponíveis para determinar doenças cardíacas com sucesso, mas de todos os algoritmos apresentados, o “Support Vector Machine” e a Rede Neuronal Artificial apresentam maior taxa de sucesso na detecção de doenças cardíacas, segundo a análise do estado da arte realizada no segundo capítulo.

Por isso, pretende-se desenvolver estes dois algoritmos e concluir qual será o algoritmo ideal na identificação do coração anormal.

3.7.3.1 “Support Vector Machine”

“Support Vector Machine” é um método de “Machine Learning” supervisionado muito eficaz que toma dados de entrada e os classifica entre várias categorias. Para este trabalho, pretende-se criar uma “Support Vector Machine” de mais do que duas entradas e que retorne uma saída única que classifique os dados como pertencente a uma entre cinco categorias. A entrada presentes neste método são os dados do ECO ou SCG. Cada categoria presente na saída deste método representa uma classe. No caso da dissertação, as classes correspondem quatro tipos de estado do coração, MI, AS, HCM, DCM e “Healthy”.

Uma “Support Vector Machine” pode ser implementada pelo sistema. O sistema pode, através dos dados oferecidos, construir uma máquina e aprender com a mesma.

O algoritmo lida com problemas não lineares mapeando o conjunto de dados de seu espaço original, referenciado como de entradas, para um novo espaço de maior dimensão e isso faz com que o conjunto de dados mapeados em novo espaço passa ser separado por uma “Support Vector Machine” linear. Isso é, supondo um espaço de entradas S de duas dimensões, transforma S para um novo espaço de três dimensões, assim, o conjunto de dados não linear torna-se linearmente separável em novo espaço. Para que seja possível realizar o processo, é necessário utilizar função “kernel”.

3.7.3.1.1 Função “Kernel” A função “kernel” é uma função que recebe dois pontos x_i e x_j no espaço de entradas e calcula o produto escalar desses objetos no espaço de características. Alguns dos “kernels” mais utilizados na prática são os polinomiais, os de função base radial (“Radial Basis Function” - RBF) e os sigmóides. Desta forma, no contexto desta dissertação, função base radial destina-se a treinar o modelo de “Machine Learning”. Tem-se então:

$$K(x_i, x_j) = \exp(-\sigma ||x_i - x_j||^2) \quad (6)$$

6: Equação de RBF

A obtenção de um classificador por meio do uso deste método envolve a escolha de parâmetro σ da Equação 6, uma vez que este parâmetro afeta o desempenho do classificador obtido, pois ele define a fronteira de estados do coração.

3.7.3.1.2 Função “loss” A Equação 7 apresenta a função “loss” que vai ser usada para determinar os estados do coração. Como já foi indicado anteriormente, quando menor for o valor de loss, maior é precisão do modelo de “Machine Learning”.

$$Loss = \sum_{s=1}^n - \left(\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) \right) \quad (7)$$

7: Equação de “loss”

3.7.3.2 Redes neurais artificiais

A rede neuronal artificial é um algoritmo de “Machine Learning” supervisionado muito conhecido e poderoso. Como já foi indicado no estado da arte, caracteriza-se por ter uma estrutura semelhante a uma rede neuronal biológica e são organizadas em três camadas: Camada de entrada onde os atributos são apresentados à rede; Camada escondida onde trata o processamento de dados; e camada de saída onde o resultado é concluído. No contexto desta dissertação, os atributos presentes na dissertação são dados da SCG ou dados do ECO. Cada processamento de dados corresponde um treino da aprendizagem e cada resultado representa uma classe (resultado obtido depois de aprender).

O modelo de redes neurais possui alguma regra de treino, onde os pesos de sua conexão são ajustados de acordo com os padrões apresentados. No caso desta dissertação, utiliza “Deep Learning” e “Feed-forward” para treinar redes neurais.

3.7.3.2.1 “Deep Learning” “Deep Learning” é uma abordagem para treinar redes neurais que se tem tornado muito popular nos últimos anos. Como o próprio nome diz, é utilizado para o treino de redes neurais profundas. Este tipo de redes é aquele que possuem pelo menos duas camadas escondidas, assim extraem dos dados de entrada características que são consideradas relevantes para extração de modelos com alta capacidade preditiva. Cada camada transforma as características extraídas pela camada anterior por meio de um processamento simples e não linear.

3.7.3.2.2 “Feedforward” “Feedforward” é um processo de treino e iterativo, constituído por apenas uma fase para frente(“forward”). Nesta fase, cada atributo de entrada é apresentado à rede. O atributo é recebido por cada um dos neurónios da primeira camada escondida da rede, sendo ponderado pelo peso

associado a suas conexões de entradas correspondentes. Cada neurónio nessa camada produz um valor de saída, que vai ser utilizado como valor de entrada pelo neurónio da camada seguinte. Este processo continua até que as redes neuronais produzem uma saída. Em fim, esta saída vai ser comparada ao valor desejado para a saída desse último neurónio. A diferença entre os valores de saída obtidos e desejados indica o erro cometido pela rede para o objeto apresentado.

4 Especificação do Sistema

Neste capítulo será ilustrado e explicitado como será detalhado o hardware, software e os dados a ser utilizados para desenvolvimento da dissertação. No que diz respeito à implementação do sistema, são apresentados o elemento de Hardware que garante o envio dos sinais da SCG, a criação de uma aplicação para analisar dados, assim como o software e os atributos disponíveis que SCG e ECO podem oferecer.

4.1 Hardware

4.1.1 Sensor SCG



Figura 22: Dispositivo de medição da SCG.

A Figura 22 representa o dispositivo de medição da SCG que vai ser utilizado neste projeto, é um sistema baseado em torno de acelerómetros MEMS que tiram proveito do efeito de “pull-in” e com características que permitem medir acelerações com elevada sensibilidade. O sensor mede o comportamento mecânico do coração e é capaz de avaliar de forma discreta os sinais fisiológicos cardíacos e respiratórios, sem necessidade de fixação de dispositivos no corpo.

O dispositivo é dividido em três camadas: camada de leitura, camada de controlo e camada de alimentação. A camada de alimentação é responsável pelo estabelecimento de toda a alimentação necessária aos componentes que integram o sistema. A camada de controlo é responsável pelo envio de dados medidos sem fio. A camada de leitura é responsável pela medição ou deteção dos sinais proveniente do comportamento do coração.

Este sistema inclui num “chip” um processador ARM Cortex-M0 de 32 bits que opera a uma frequência de 48 MHz. Além disso, possui um módulo Bluetooth (BLE - “Bluetooth Low Energy”), e ainda blocos de

hardware analógico e digitais programáveis.

4.2 Software

4.2.1 Linguagem de programação

As linguagens de programação a serem usadas nesta dissertação estão presentes a seguir. A linguagem C/C++ irá ser necessária para o desenvolvimento do respetivo código de envio do sinal obtido do dispositivo da SCG e a linguagem “Python” irá ser utilizada na implementação de aplicação no “Personal Computer”.

- “Python”
- C/C++

4.2.2 Bibliotecas

“Python” é uma linguagem de alto nível que suporta várias bibliotecas capazes de oferecer uma grande variedade de ferramentas para analisar os dados. A seguir, apresentam-se algumas bibliotecas que podem ser utilizada na dissertação.

4.2.2.1 “numpy”

“numpy” é uma biblioteca de processamento de matriz, o seu principal objetivo em relação à análise de dados é armazenar e transmitir os dados entre os algoritmos. Além disso, “numpy” oferece muitas ferramentas avançadas de programação numéricas, como processamento vetorial e aritméticas sofisticadas. “numpy” é uma maneira muito eficiente de manipular dados do que as outras estruturas de dados do “Python”.

4.2.2.2 “scipy”

“scipy” é uma biblioteca que oferece várias ferramentas de algoritmos científicos, sobretudo, o algoritmo de processamento de sinais, a biblioteca suporta algumas funções de filtragem como passa-banda.

4.2.2.3 “matplotlib”

“matplotlib” é uma biblioteca de plotagem 2D em “Python”. As ferramentas da biblioteca permitem converter dados em um estado visível como gráfico, histograma e gráfico de dispersão.

4.2.2.4 “Tensorflow”

“Tensorflow” é uma biblioteca de software de “Open Source” para computação numérica. Como já foi mencionado no estado da arte, ele funciona como um gráfico de fluxo, nos quais cada nó do gráfico representa uma operação matemática. A biblioteca suporta várias ferramentas de algoritmos para implementar e treinar o modelo de “Machine Learning”.

4.2.2.5 csv

csv é uma biblioteca que utiliza estruturação específica para organizar dados. A biblioteca oferece ferramentas para ler e armazenar os dados em arquivo.

4.2.2.6 “seaborn”

“seaborn” é uma biblioteca de visualização de dados “Python” baseada no “matplotlib”. Ele fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos.

4.2.2.7 “pandas”

“pandas” é uma biblioteca que fornece estrutura de dados de alta desempenho e ferramentas avançadas, é projetada para tornar o trabalho com dados estruturados rápidos, fácil e expressivo. A biblioteca fornece funcionalidade de indexação sofisticada para facilitar a reformulação de dados, a execução de agregações.

4.2.2.8 pywt

pywt (“Wavelet Transforms in Python”) é uma biblioteca de software de transformação “Wavelet” de “Open Source”. Ele combina uma interface simples de alto nível com um desempenho de baixo nível C e “Cpython”. A biblioteca suporta vários algoritmos como cálculo reais e complexo, cálculos de precisão simples e Transformada “Wavelet” discreta.

4.3 Ambiente integrado de desenvolvimento de programas

4.3.1 PyCharm

Pycharm é um ambiente de desenvolvimento usado na programação de computadores, especificamente para a linguagem “Python”, que permite recursos avançados para o desenvolvimento de aplicativos. Além de oferecer a “Open Source TensorFlow”, suporta várias bibliotecas úteis para desenvolver o trabalho pretendido.

4.3.2 Psoc Creator

O Psoc Creator é um ambiente de desenvolvimento que permite a edição simultânea de hardware e “firmware”. As aplicações são implementadas usando captura esquemática. Oferece ferramenta de criação fácil de usar, suporta biblioteca de comunicação completa como “Bluetooth Low Energy” e leitura do batimento cardíaco.

4.4 Dados disponíveis

Esta secção pretende-se apresentar os dados que vão ser utilizados no ambiente desta dissertação. Como já foi mencionado anteriormente, os dados do ECO e da SCG têm vantagem em relação aos outros tipos de dados, por isso vão ser escolhidos.

4.4.1 Dados provenientes da SCG

Em seguida apresenta as ondas típicas da SCG, como já foi indicado explícito no estado da arte, é necessário realizar um método para extrair os parâmetros (tais como frequência cardíaca) a partir destas ondas oferecidas e com base nestas ondas podem oferecer muitos parâmetros. Em baixo, pretende-se mostrar essas ondas mais em detalhe.

- AS (“Atrial systole”)
- IM (“Isovolumic Moment”)
- AO (“Aortic Opening”)
- IC (“Isovolumic Contraction”)
- RE (“Rapid Ejection”)
- AC (“Aortic Closure”)
- ACM (“Diastolic Periods”)
- MO (“Miral Opening”)

4.4.2 Dados provenientes do ECO

Embaixo apresenta os dados do ECO, como pode ser visto que os dados do ECO apresentam mais explícitos do que os dados da SCG, uma vez que o dispositivo do ECO oferece diretamente as informações úteis sobre o comportamento do coração.

- HR ("Hearttrate") [$b.p.m$]
- SV ("Systolic Volume") [mL]
- PEP ("Pre-ejection Period") [$m.s$]
- EF ("Ejection Fraction") [%]
- IVSd ("Interventricular Sepetal thinckness") [mm]
- PP [mm]
- LVMass ("Left Ventricle Mass") [g]
- LVM ("Left Ventricle index") [g/m^2]
- EDV ("End-diastolic volume") [mL]
- LVED ("End-diastolic diameter") [mL]
- ET ("Ejection Time") [$m.s$]
- MFR ("Maximum Flow Rate") [m/s]

5 Implementação do Sistema

5.1 Fluxograma

A Figura 23 mostra o fluxograma da inicialização do sistema. De acordo com a figura apresentada, é possível observar a inicialização de todos os componentes do sistema e ainda algoritmos de identificação de sinais da SCG e determinação de doenças cardíacas através da técnica de “Machine Learning”. Para algum componente e para os algoritmos foram criadas umas sub-rotinas de inicialização para que o código seja melhor estruturado. Os fluxogramas associados a estas sub-rotinas são apresentados a seguir.

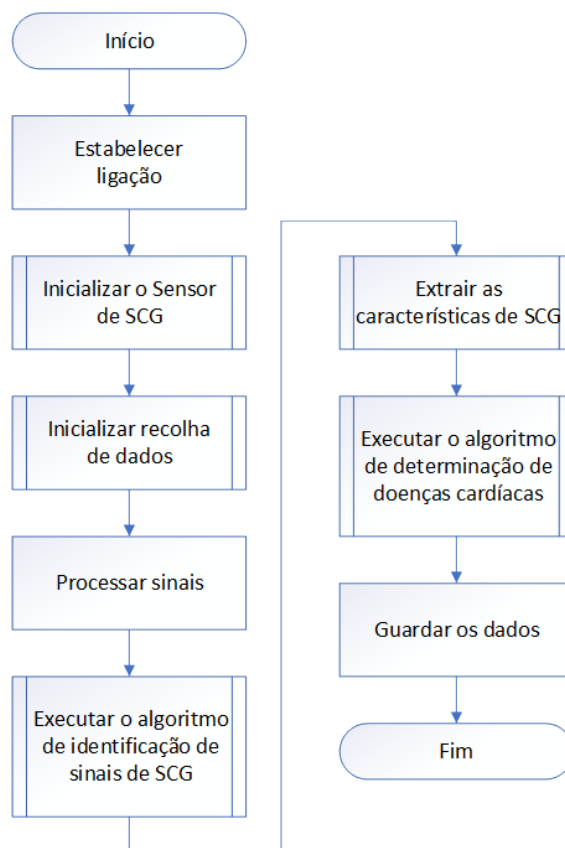
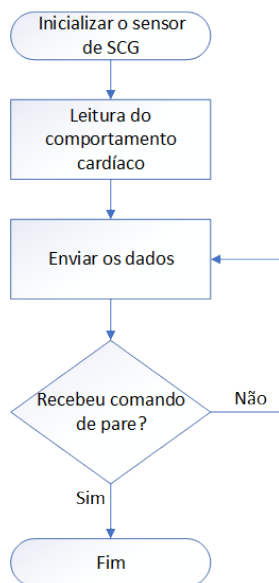
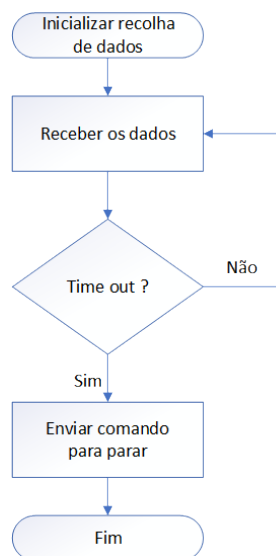
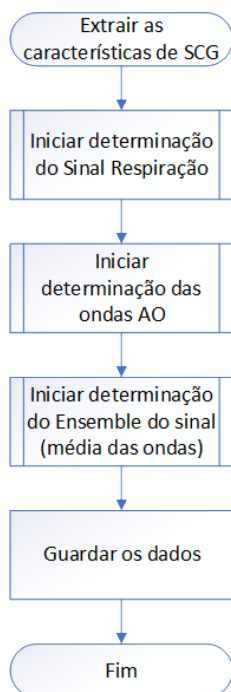


Figura 23: Fluxograma do sistema.

**Figura 24:** Fluxograma do envio**Figura 25:** Fluxograma da recepção

As Figuras 24 e 25 mostram os fluxogramas de envio dos dados e recolha dos dados. Estas duas tarefas apenas são executadas durante a comunicação estabelecida entre o dispositivo do sensor e a aplicação. Quando a comunicação é estabelecida, os dados serão transmitidos através do “Bluetooth” durante tempo definido, após este tempo, a aplicação enviará um comando para o sensor parar de enviar.

**Figura 26:** Fluxograma do sistema.

A Figura 26 ilustra o processo de extração das características da SCG. Esta tarefa é responsável, tal como

foi referido no capítulo 3, pela aquisição e cálculo das informações necessárias a partir de sinal da SCG. O fluxograma apresentado na figura demonstra passo a passo a inicialização da aquisição ou extração das características do sinal da SCG, utilizando o método referido no capítulo 3. Cada inicialização vai ser apresentado em mais detalhe a seguir. Por fim, as características obtidas serão guardadas no armazenamento de dados.

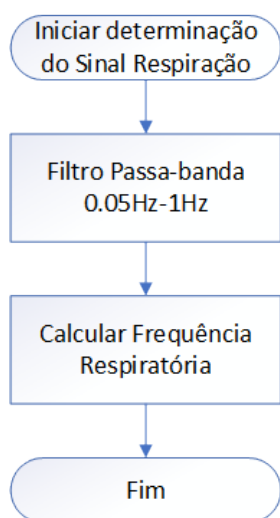


Figura 27: Fluxograma da determinação de frequência respiratória

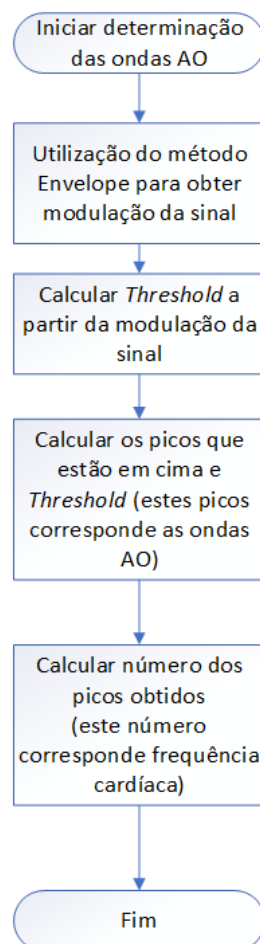


Figura 28: Fluxograma da determinação de frequência cardíaca

As Figuras 27 e 28 apresentam os processos para determinar a frequência respiratória e frequência cardíaca. A partir dos dados lidos será calculado o valor da frequência respiratória utilizando o método do filtro passa-banda para a gama de frequências [0.05Hz – 1Hz]. De seguida é calculado o valor da frequência cardíaca, vai ser aplicando o método Envelope para obter modulação do sinal, e depois determinar o valor de “threshold” para determinar os picos que estão em cima desta “threshold” calculado e quantidade dos picos obtidos representam a frequência cardíaca. .

Na Figura 29 encontra-se representado o fluxograma da determinação da média das ondas. Este processo é responsável pelo cálculo do respetivo valor de amplitudes de cada onda e o intervalo de tempo entre

elas a partir da média das ondas determinada. Para determinar essa média das ondas, é necessária realizar uma sobreposição dos vários intervalos identificados pelo pico AO.

Na Figura 30 está descrito o processo da determinação de correlações entre ECO e SCG. Esta tarefa apenas é chamada depois de ter calculado as características a partir de sinal de SCG. Para que seja possível uma observação, torna-se necessário o processo de criação dos gráficos para visualizar os dados em duas dimensões. A seguir, utiliza o método “Pearson” para determinar alguma possível correlação entre ECO e SCG.

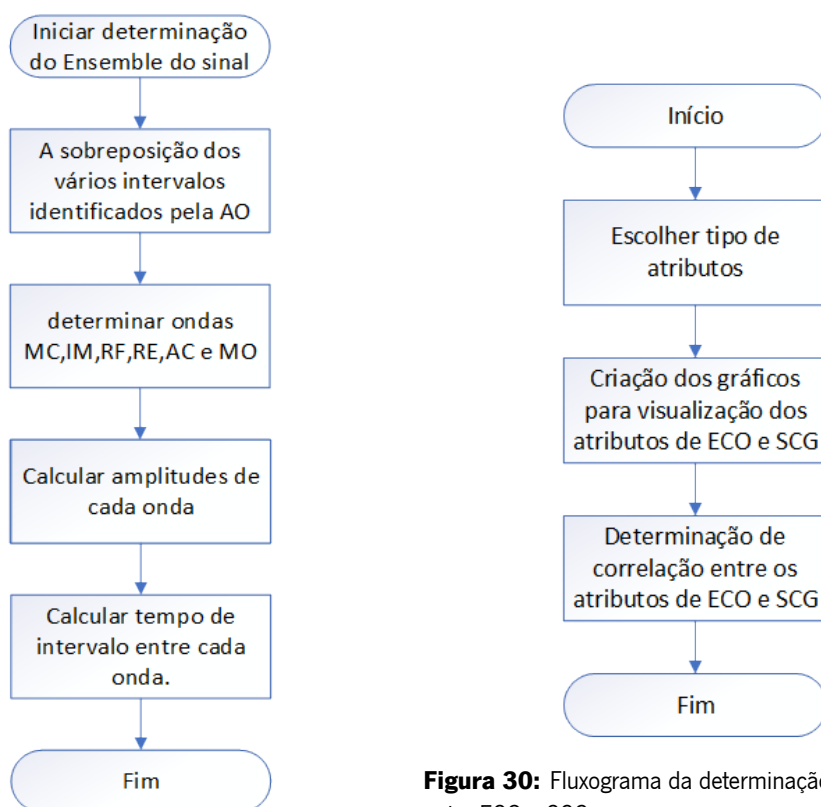


Figura 29: Fluxograma da determinação da onda médio da sinal da SCG

Figura 30: Fluxograma da determinação de correlação entre ECO e SCG

5.2 Algoritmos de “Machine Learning”

A seguir será apresentado como serão feitas as estruturas dos algoritmos e ao detalhe o funcionamento de cada algoritmo. Para isso optou-se por o uso de pseudocódigo, visto que explicam os algoritmos de “Machine Learning” mais detalhadamente.

5.2.1 “K-Means clustering”

O Algoritmo 1 é representado a estrutura de “K-Mean”. Este algoritmo recebe um conjunto de dados e agrupa automaticamente os dados. Como já foi mencionado anteriormente, ele é um algoritmo de “Machine Learning” não supervisionado, ou seja, não existe fase de treino.

No início, é necessário calcular desvio padrão e média a partir do sinal proveniente do sensor SCG e atribui um valor ao K. Depois, gera k centroides aleatórios e calcula a distância do dado ao centroide de cada k “clusters”. A seguir, o dado vai ficar no “cluster” que tiver a menor valor da distância. Por fim, realiza um ciclo para recalculer os novos valores de k “clusters” e reagrupar os dados de acordo com a distância ao novo centroide de cada k “clusters”, o ciclo só acaba quando os dados forem totalmente agrupados.

Algorithm 1 “K-Mean”

Entrada: Um conjunto de dados

Saída: cluster atribuída aos dados

início

 Calcular desvio padrão : s;

 Calcular média : m;

 Preparar conjunto de dados $\{s, m\}$;

 Informar o valor de K;

fim

para cada cluster K **faça**

 Atribuir valor ao centroide de cada cluster;

enquanto critério de parada não for atingido **faça**

para cada dado **faça**

 Calcular a distância do dado ao centroide de cada cluster ;

 O dado irá ficar no cluster que tiver a menor distância;

para cada cluster K **faça**

 └ Recalculer os centroides.

retorna conjunto de K centroides;

5.2.1.1 Realização dos testes sobre “K-Mean”

Depois de ter o modelo feito, é necessário testar este modelo. O Algoritmo 2 apresenta o comportamento do teste.

Algorithm 2 Teste

Função teste(conjunto de K centroides)

 Executar o algoritmo para agrupar os novos dados

Fim

5.2.2 Regressão logística múltipla

O Algoritmo 3 mostra o pseudocódigo do treino baseado em regressão logística múltipla. O algoritmo recebe um conjunto de dados de treino e constrói um modelo para identificar sinais da SCG. Como é possível observar no algoritmo, cada treino é necessário executar as funções “softmax_logits” e “softmax_accurate” e em seguida, introduz essas duas funções em mais detalhes.

Algorithm 3 Regressão logística múltipla

Entrada: Um conjunto de dados

Saída: modelo treinado

início

 Preparar conjunto de dados $\{desvio_{padro} : v, mdia : m\}$;

fim

para cada treino(v, m) **faça**

 Executar softmax_modelo(v,m);

 Executar softmax_accurate(v,m);

retorna um conjunto de distribuições;

5.2.2.1 “softmax” e “loss”

O Algoritmo 4 ilustra a função softmax_logits. Esta função é responsável pela criação do modelo de “Machine Learning”, utilizando o método regressão “softmax” e “cross_entropy”. Além disso, a função calcula “loss” e utiliza técnica de otimização oferecida pela “Tensorflow” para minimizar essa “loss”.

Algorithm 4 “Softmax”

Função softmax_modelo(v,m)

 atualizar o coeficiente a;

 atualizar o coeficiente b;

 Calcular $y=av+b$;

 Calcular $y1 = \frac{\exp(y_j)}{\sum_{j=1}^k \exp(y_j)}$;

 Calcular $H=m \times \log(y1)$;

 Calcular loss= mean(H);

 Minimizar loss;

Fim

5.2.2.2 Realização da avaliação sobre o modelo da regressão logística múltipla

O Algoritmo 5 representa a função “softmax_accurate”. A função é responsável pela avaliação do modelo obtido após cada treino. Uma vez que é necessário confirmar a sua “accuracy”.

Algorithm 5 “Accuracy”

Função softmax_accurate(v,m)
 Comparar o resultado previsto com o resultado real;
 Calcular a accuracy do modelo;
Fim

5.2.2.3 Realização dos testes sobre o modelo da regressão logística múltipla

Depois de ter treinado o modelo, é necessário implementar um algoritmo de teste para validar o modelo.

O Algoritmo 6 é apresentado o comportamento do teste.

Algorithm 6 Teste

Função teste(novo dado)
 Executar o algoritmo softmax e retornar a probabilidade de que
 os dados de entrada pertençam as distribuições determinadas;
Fim

5.2.3 SVM

O Algoritmo 7 representa o pseudocódigo do treino baseado em SVM. O algoritmo recebe um conjunto de dados de treino e constrói um modelo para classificar o estado do coração. Como é possível observar no algoritmo, cada treino é construído pelas funções “SMV_modelo” e “SMV_accurate” e estas duas funções serão ser explicadas mais em detalhe em seguida.

Algorithm 7 SVM

Entrada: Um conjunto de dados de treino
Saída: SVM treinado
início
 | Preparar conjunto de dados $\{X, Y\}$;
 | Iniciar a variável α ;
 | Definir o parâmetro σ
fim
para cada treino(X, Y) **faça**
 | Executar SMV_modelo(X,Y);
 | Executar SMV_accurate(X,Y);
retorna Classificador

O Algoritmo 8 mostra a função “SMV_modelo”. A função é responsável pela geração do modelo da previsão, utilizando o método “Radial Basis Function” (RBF). Depois é calculado o valor “loss” e este valor vai ser minimizado pela técnica da otimização que é oferecida pela “Tensorflow”.

Algorithm 8 RBF**Função** SVM_modelo(X,Y)Atualizar α ;Calcular dimensão= $2 \times X \times X^T$;Calcular RBF= $\exp(\sigma * |distancia|)$;Calcular $f = \sum \alpha - \frac{1}{2} \sum \sum Y Y^T \alpha \alpha^T RBF$;Calcular loss= $sum(-f)$;

Minimizar loss;

Fim**5.2.3.1 Realização da avaliação sobre o modelo SVM**

O Algoritmo 9 representa a função “SMV_accurate”. A função é responsável pelo cálculo da “accuracy” do modelo criado após cada treino.

Algorithm 9 “Accuracy”**Função** SMV_accurate(X,Y)

Comparar o resultado previsto com o resultado real;

Calcular a accuracy do modelo;

Fim**5.2.3.2 Realização dos testes sobre o modelo SVM**

Depois de ter modelo da previsão, é necessário realizar um teste, o Algoritmo 10 ilustra o comportamento do teste.

Algorithm 10 Teste**Função** teste(entrada);

Correr classificador e indicar as classes que os dados devem pertencer;

Fim**5.2.4 Rede Neuronal da DNN**

O Algoritmo 11 representa a estrutura do treino da rede neural da “Deep Neural Networks” (DNN), utilizando a técnica de “forward propagation”. De acordo com o pseudocódigo mostrado, antes de iniciar o treino é necessário definir bem os parâmetros pretendidos para este algoritmo e depois será executada uma função para avaliar o modelo treinado e a sua “accuracy” associada.

Este algoritmo será implementado com facilidade, pois a biblioteca “TensorFlow” oferece bastantes ferramentas da DNN para realizar o algoritmo.

Algorithm 11 Redes neuronais da DNN

Entrada: Um conjunto de dados de treino

Saída: Avaliação de modelo

início

- Preparar um conjunto de dado de treino;
- Preparar um conjunto de dados de teste;
- Definir número de camadas;
- Definir unidade neuronal;
- Definir número de classes desejadas;
- Treinar o modelo;
- Avaliar o modelo;
- prever novos conjuntos de dados;

fim

6 Resultados obtidos

Neste capítulo serão apresentados os resultados obtidos no decorrer desta dissertação e dividem-se em varias partes. Primeiramente é realizada o filtro do sinal proveniente do dispositivo da mediação da SCG. De seguida, é feita a identificação de sinais da SCG através do algoritmo de "Machine Learning" e uma consequente análise dos dados. A análise dos dados consiste na determinação dos parâmetros mencionados na fase de análise do sistema a partir do sinal da SCG, que visam a determinação de eventos cardíacos. Por fim é feita a determinação da patologia cardíaca.

6.1 Resultado do sinal recebido

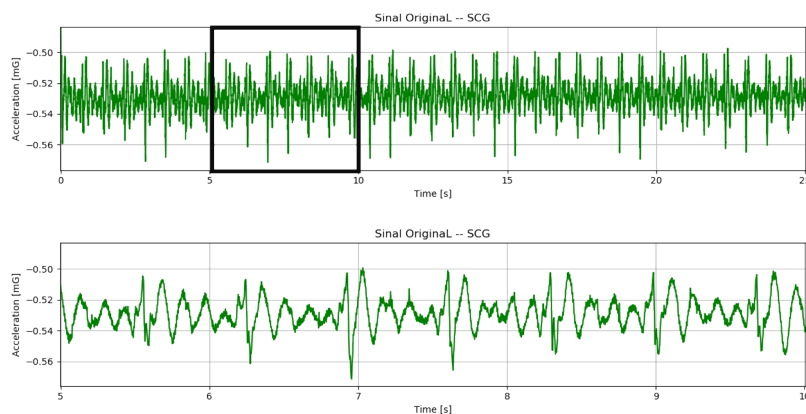


Figura 31: Representação gráfica do sinal obtido pelo dispositivo da SCG.

A Figura 31 apresenta uma amostra de 5 segundos do sinal da SCG original adquirido pelo dispositivo. Pela análise da figura, é possível verificar que o sinal da SCG apresenta uma componente com baixa frequência. Tal se verifica devido ao movimento provocado pela respiração do paciente submetido à análise. Os sinais são transferidos via BLE para a aplicação, que o exibe na interface, permitindo ao utilizador analisar os resultados obtidos.

6.2 Resultado dos filtros de Sinais

Com o intuito de remover a componente DC associada à aceleração da gravidade e as componentes de baixa frequência proveniente da respiração, o presente trabalho propõe dois métodos de filtro capaz de filtrar o sinal proveniente do dispositivo da SCG: Passa-banda e Transformada Discreta "Wavelet". Segundo a análise da metodologia realizada no terceiro capítulo, tanto um como outro, ambos os métodos são eficazes para remover o componente indesejado que existe no sinal da SCG.

6.2.1 Passa-banda

Primeiro, foi aplicado ao sinal da SCG um filtro passa-banda para o intervalo de 1 Hz – 20 Hz, pois a informação relativa ao sinal da SCG integra o intervalo de frequências de 1 Hz – 10 Hz, e para que fosse possível avaliar a melhor este método, foram extraídos os sinais da SCG provenientes de cinco tipos diferentes de pacientes: DCM, HCM, MI, AS e Healthy. Os resultados obtidos podem ser visualizados na Figura 32, 33, 34, 35 e 36.

As Figuras 32, 33, 34, 35 e 36 apresentam uns exames de 60 segundos do sinal da SCG. Para melhor visualização deste sinal, o lado direito dos exames apresentam também uma amostra de 10 segundos.

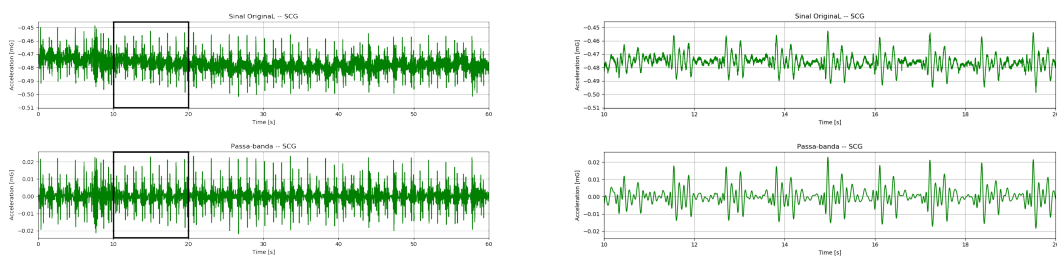


Figura 32: Sinal proveniente do paciente DCM

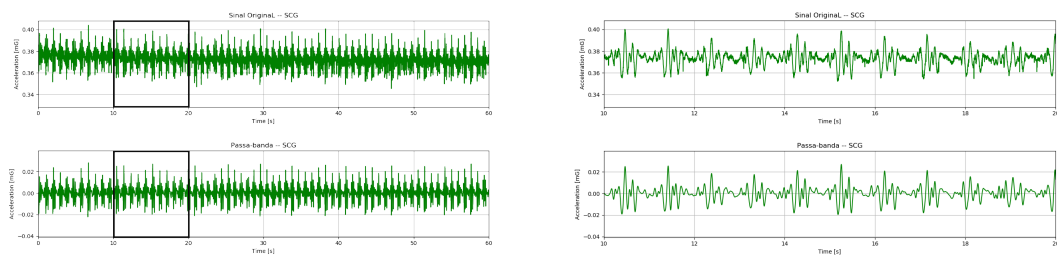


Figura 33: Sinal proveniente do paciente HCM

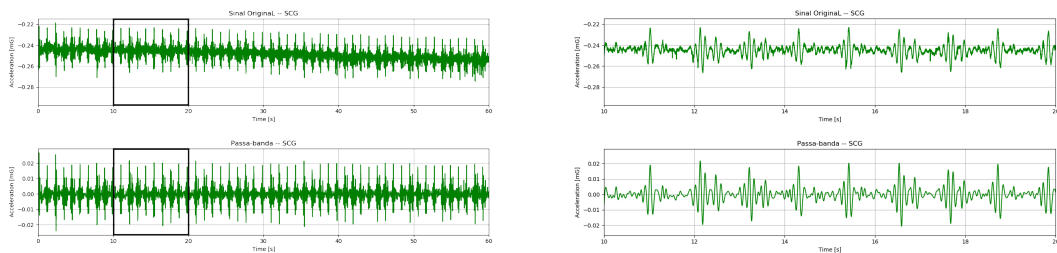


Figura 34: Sinal proveniente do paciente MI

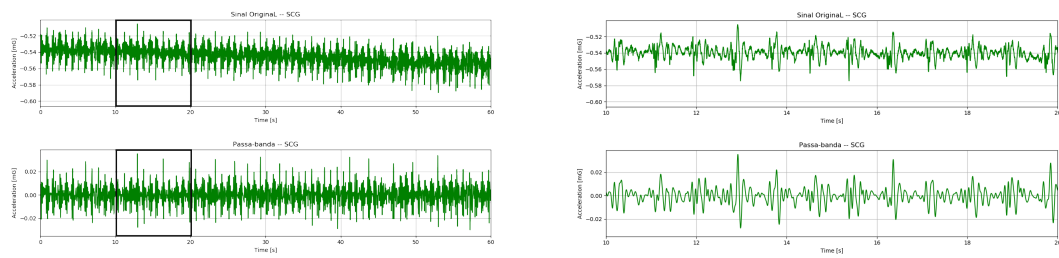


Figura 35: Sinal proveniente do paciente AS

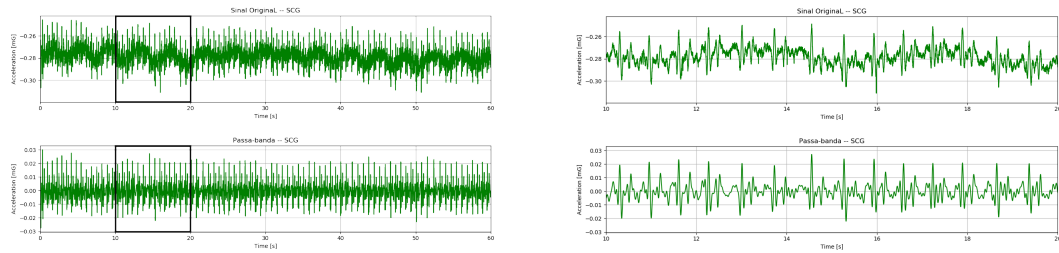


Figura 36: Sinal proveniente do paciente Healthy

De acordo com os resultados preliminares da extração de sinais é possível verificar que este método conseguiu eliminar os componentes com baixa frequência proveniente da respiração e componente DC associado à aceleração da gravidade, e através da aplicação deste método pode determinar o sinal da SCG. Além disso, com este método é possível identificar melhor o ritmo do batimento cardíaco e os picos de maior amplitude para cada pulso.

6.2.2 Transformada Discreta Wavelet

O método de Transformada Discreta "Wavelet" foi realizado com os mesmos dados utilizados no teste anterior. Antes disso, existem dois fatores principais que têm impacto direto no desenvolvimento deste método: número de camada de decomposição e famílias de "Wavelet". Este teste utilizou "Symlet".

Para poder determinar o número ideal de camadas de decomposição, neste teste foram utilizadas 2 até 7 camadas.

Analisando a Figura 37, é possível verificar que o gráfico 4 apresenta o melhor efeito e está mais próxima do sinal original. De acordo com o resultado, os gráficos 1, 2 e 3 apresentam ainda alguns picos indesejados, o que significa que o número da decomposição de camada não foi suficiente para remover algum componente com baixa frequência. Os gráficos 5, 6 e 7, apesar de ter sucesso na eliminação dos componentes indesejados, estão distorcidos, sobretudo nos gráficos 6 e 7, onde ocorreram distorções graves.

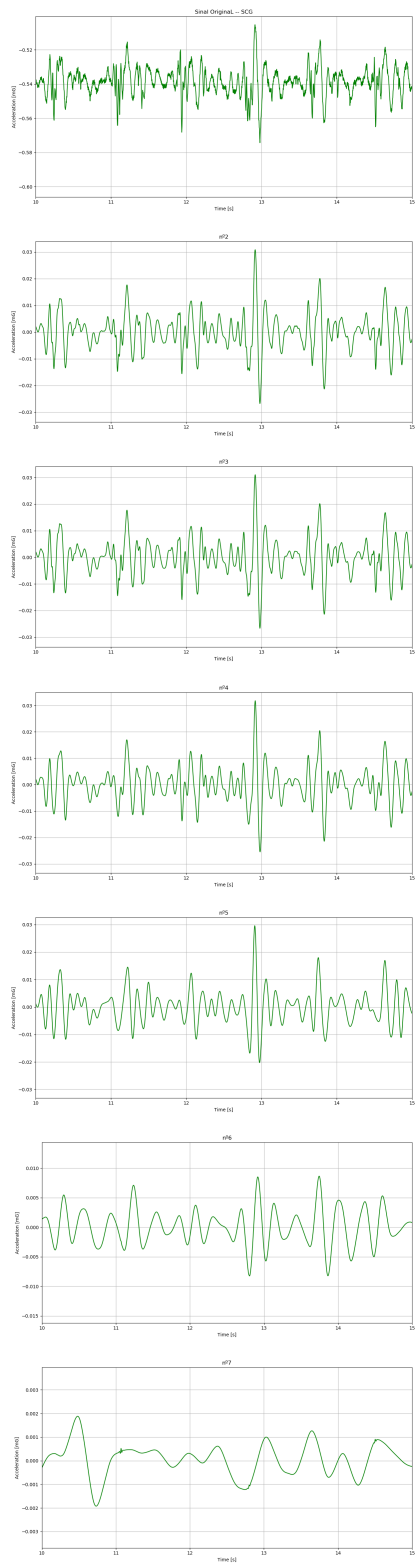


Figura 37: Resultados da decomposição

Depois de ser determinado o número da decomposição das camadas e "Wavelet", é necessário realizar um teste para que fosse possível avaliar a melhor método. Os resultados obtidos podem ser observados nas Figuras 38, 39, 40, 41, e 42.

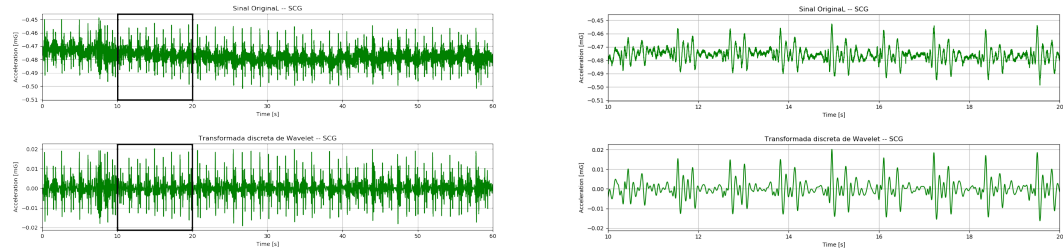


Figura 38: Sinal proveniente do paciente DCM

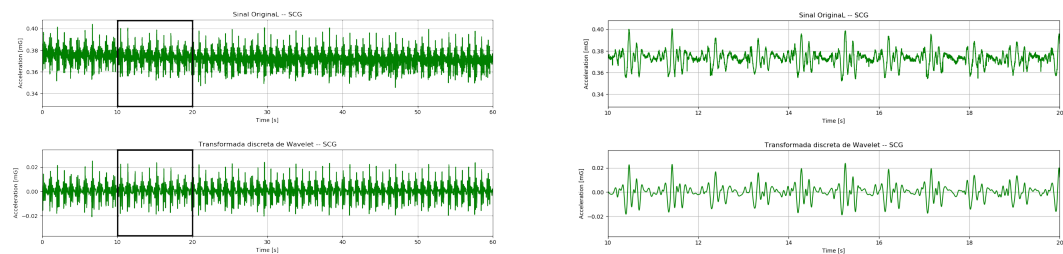


Figura 39: Sinal proveniente do paciente HCM

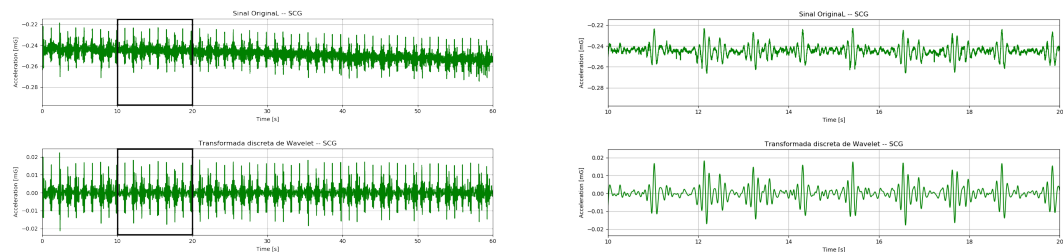


Figura 40: Sinal proveniente do paciente MI

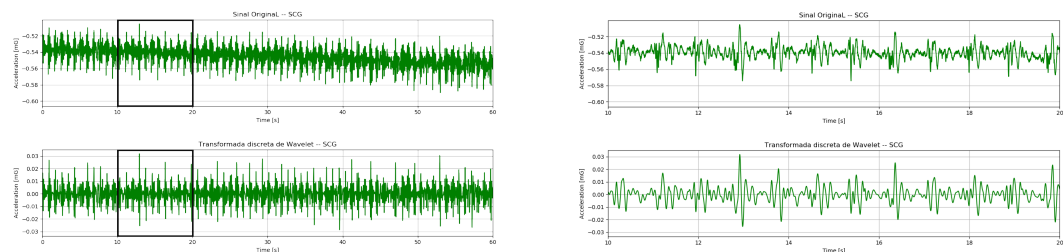


Figura 41: Sinal proveniente do paciente AS

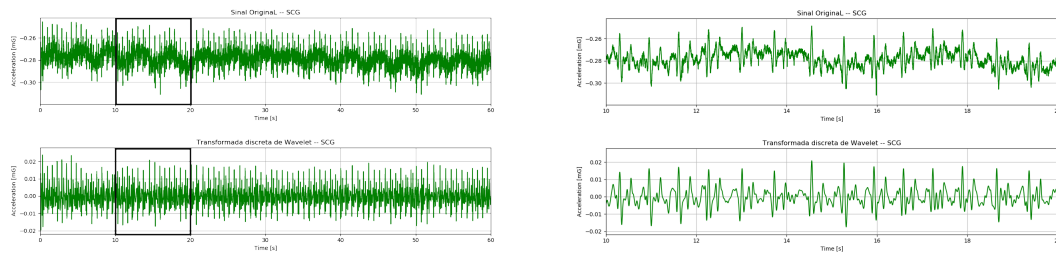


Figura 42: Sinal proveniente do paciente Healthy

Segundo os resultados representados nas figuras é possível ver que o método aplicado teve sucesso na eliminação dos componentes indesejados, consequentemente permitindo obter um sinal de melhor qualidade. Além disso, o método permite já identificar algumas características de sinais da SCG, tais como o ritmo do batimento cardíaco e os picos de maior amplitude para cada pulso.

6.2.3 Discussão do resultado

Analisando os resultados obtidos nos dois métodos, é possível concluir que ambos apresentaram bons resultados, mas o método Transformada discreta "Wavelet" obteve melhores resultados que o passa-banda. Tal pode ser observado nas Figuras 43, 44, 45, 46 e 47.

Para que fosse possível concluir o melhor método entre estes dois métodos, foram utilizados e comparados os resultados obtidos nos testes anteriores. Os resultados estão apresentados nas figuras, onde é possível ver a representação de sinal filtrado pelo método passa-banda, método Transformada Discreta "Wavelet" e sinal original. Para ter melhor visualização, as figuras apresentam apenas uma amostra de 4 segundos do sinal da SCG.

Pela análise, podemos verificar que os sinais obtidos pelos dois métodos são muito semelhantes, mas ainda existe alguma diferença entre eles. Como pode ser observado nas figuras embaixo, os sítios onde foram marcados a vermelho representam a distinção entre estes dois métodos.

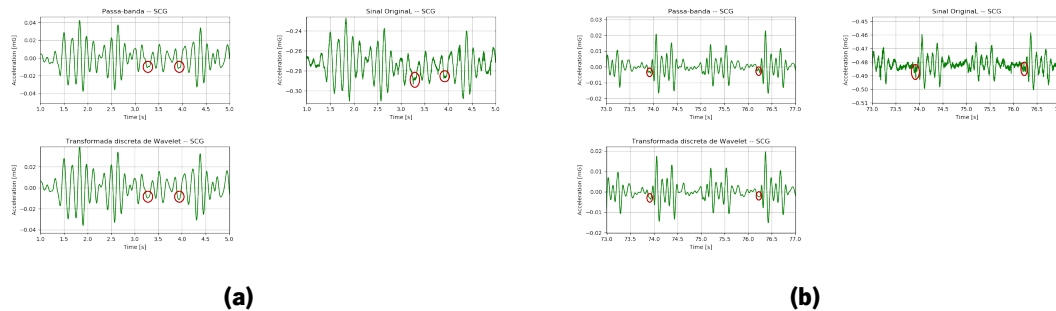


Figura 43: Sinal proveniente do paciente DCM

As Figuras 43a e 43b apresentam sinais da SCG proveniente do paciente DCM, nas quais se pode

observar primariamente duas diferenças.

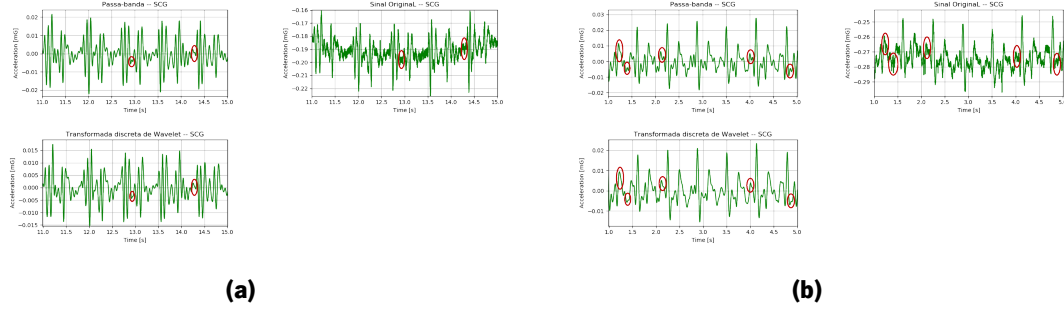


Figura 44: Sinal proveniente do paciente Healthy

As Figuras 44a e 44b apresentam sinais da SCG provenientes do paciente "Healthy". Em que a Figura 44a exibe duas diferenças entre os sinais resultantes, enquanto que a Figura 44b apresenta cinco diferenças.

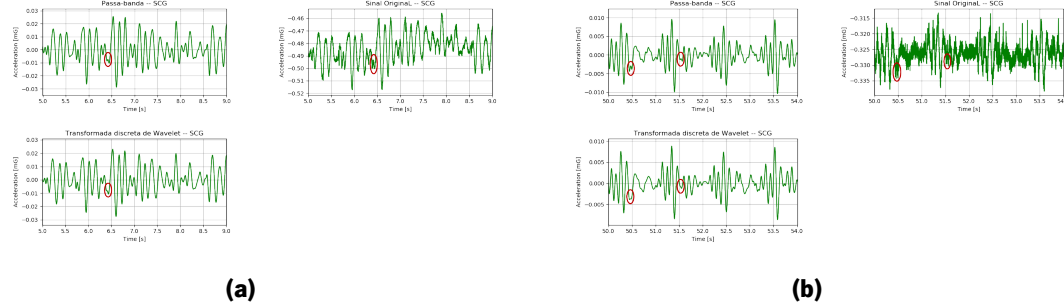


Figura 45: Sinal proveniente do paciente MI

As Figuras 45a e 45b apresentam sinais da SCG proveniente do paciente MI. Sendo que a Figura 45a demonstra uma diferença entre os sinais, por sua vez, a Figura 45b apresenta duas diferenças.

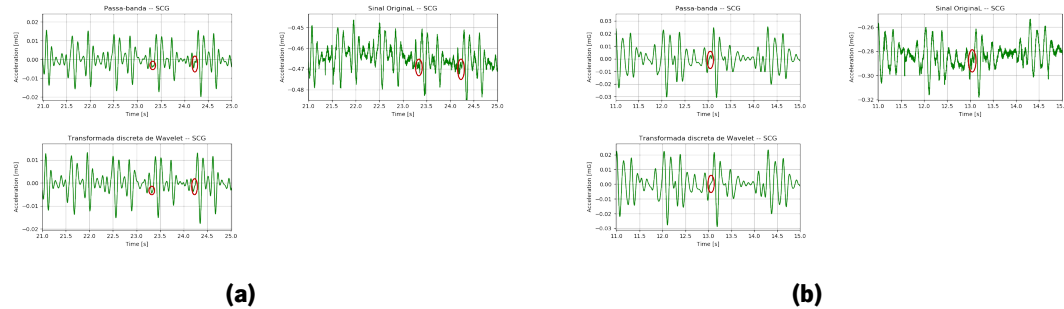


Figura 46: Sinal proveniente do paciente HCM

As Figuras 46a e 46b apresentam sinais da SCG proveniente do paciente HCM. Estando representadas duas diferenças na Figura 46a e uma diferença na Figura 46b.

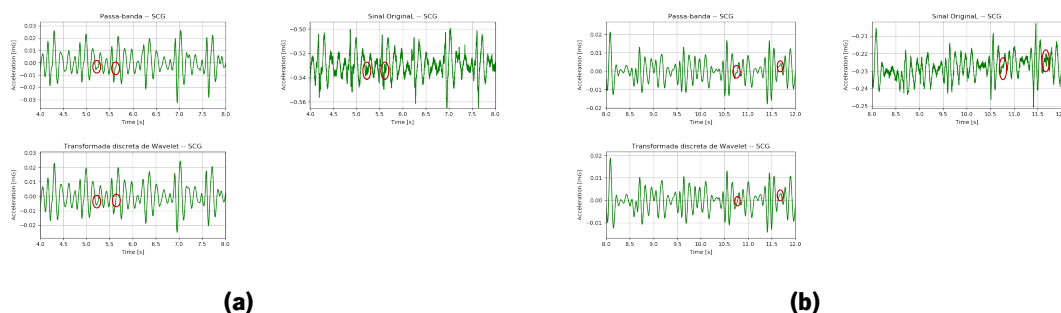


Figura 47: Sinal proveniente do paciente AS

As Figuras 47a e 47b apresentam sinais da SCG proveniente do paciente AS, ambas as figuras mostram duas diferenças entre os sinais obtidos por cada método.

As diferenças mostradas acima estão em picos. O sinal filtrado pelo passa-banda apresenta mais picos do que o sinal filtrado pela Transformada Discreta "Wavelet" e, comparado com o sinal original, esses picos podem ser considerados como ruído ou componente com baixa frequência proveniente da respiração do paciente.

Com os resultados de cima, é possível concluir que o método passa-banda não conseguiu remover alguns componentes indesejados, logo algum pico que está apresentado no sinal filtrado pelo passa-banda é considerado pico inválido ou ruído. Comparando com o sinal filtrado pela Transformada Discreta "Wavelet", o sinal já não apresenta esse tipo de picos, uma vez que o método conseguiu eliminar esses picos inválidos. Além disso, o método Transformada Discreta "Wavelet" permitiu obter um sinal mais suave do que o método passa-banda.

6.3 Resultado da identificação de sinais da SCG

Nesta secção apresentam-se os resultados obtidos aquando da avaliação dos modelos de "Machine Learning" desenvolvidos para identificar sinais da SCG. Para tal, foram realizados alguns testes a partir dos dados criados em ambiente Laboratorial e os modelos foram simulados no "Personal Computer".

Para testar a robustez do sistema, foi criada uma base de dados em que, 70% dessa base de dados foram utilizados para treinar os diferentes modelos desenvolvidos (K-Mean e regressão logística múltipla) e os restantes 30% foram utilizados para testar esses mesmos modelos, de maneira a que os dados que foram utilizados para treinar, não fossem os mesmos que foram utilizados para testar e validar os modelos.

Para fazer com que o mecanismo de "Machine Learning" aprenda melhor, os sinais foram partidos em vários segmentos. De seguida, calculou-se o valor médio e o desvio padrão a partir de cada segmento, ambos os quais representam a entrada do mecanismo de "Machine Learning".

6.3.1 Resultado do modelo K_Mean

A Figura 48 mostra os resultados do treino do modelo "K-Mean". No algoritmo usado para o modelo "K-Mean", o parâmetro k usado foi 4. Pela análise do mesmo, é possível observar que o modelo dividiu os dados em quatro cores diferentes, azul e amarelo, representam um dado válido, verde e vermelho representam dado inválido. O modelo apresenta uma "accuracy" de cerca de 99%. Além disso, na mesma figura, ao lado do gráfico, pode verificar os valores de centroides calculados, valores estes que se dividem em 4 "clustering". Os "clustering" 1 e 4 representam a classe dos dados inválidos e os "clustering" 2 e 3 representam a classe dos dados válidos. Por fim, esses "clustering" foram utilizados para identificar os dados validos e inválidos.

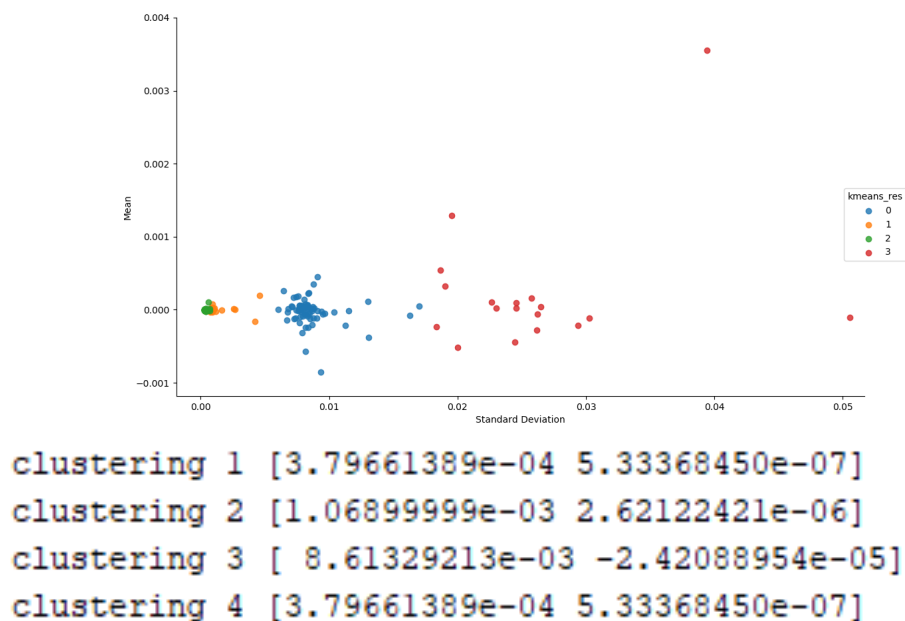


Figura 48: Resultado do modelo K-Mean

6.3.1.1 Resultados dos testes

De forma a avaliar o algoritmo "K-Mean", foi necessário realizar alguns testes. Para tal, foram recolhidos vários sinais da SCG para analisar a precisão do algoritmo. Os resultados obtidos dos diferentes testes podem ser observados nas Figuras 49, 50, 51, 52, 53 e 54, que dão uma ideia em geral dos resultados possíveis.

Analisando os resultados obtidos nas Figuras 49, 50, 51, 52 e 53, é possível observar que as figuras apresentam um resultado perfeito, pois o modelo identificou todos os sinais válidos e removeu sinais inválidos. No entanto, a Figura 54 já não apresenta um resultado esperado, uma vez que o modelo tratou o sinal válido como sinal inválido, removendo o sinal válido. Apesar de ter algumas falhas, a maioria dos testes tiveram sucesso e no fim dos testes realizados, foi possível concluir que o modelo apresentou uma precisão de acerca de 97%.

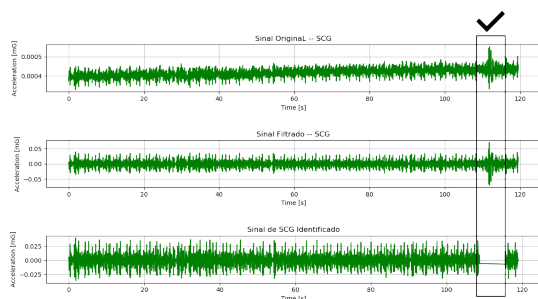


Figura 49: Teste sucedido

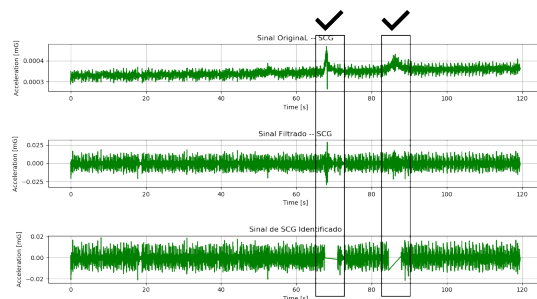


Figura 50: Teste sucedido

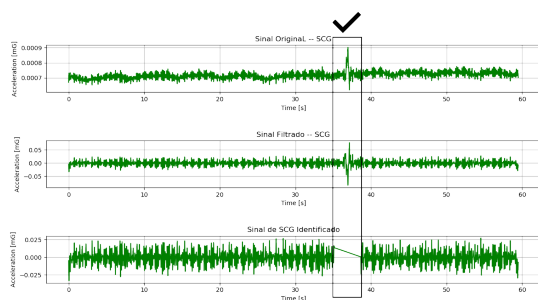


Figura 51: Teste sucedido

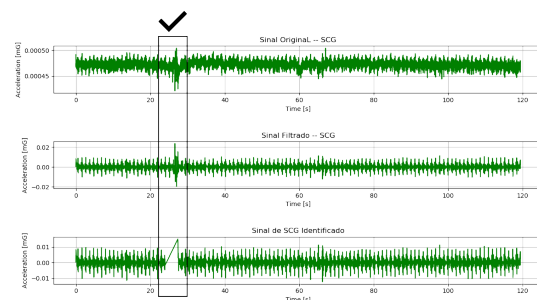


Figura 52: Teste sucedido

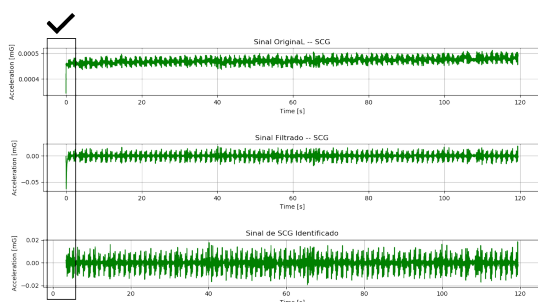


Figura 53: Teste sucedido

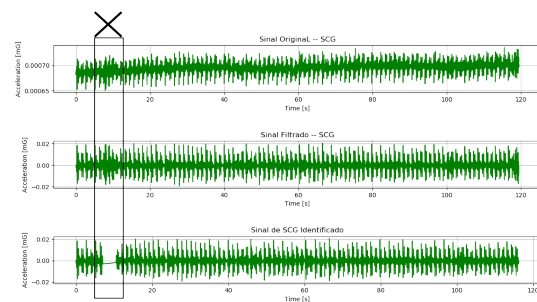


Figura 54: Teste falhado

6.3.2 Resultado do modelo regressão logística múltipla

A Figura 55 apresenta os resultados do modelo regressão logística múltipla. Analisando a figura, é possível verificar que o modelo construiu três regiões diferentes, nas quais a cada uma corresponde uma classe. A classe vermelha e castanha apresentam a área de sinais inválidos e a classe azul apresenta a área de sinais válidos. Com base nisso, o modelo atribuiu os dados recebidos a diferentes regiões. Como é possível observar, os dados apresentam três cores diferentes, os dados amarelos e vermelhos apresentam dados inválidos e os dados azuis apresentam dados válidos. Por fim, o modelo apresenta uma "accuracy" de acerca de 98%.

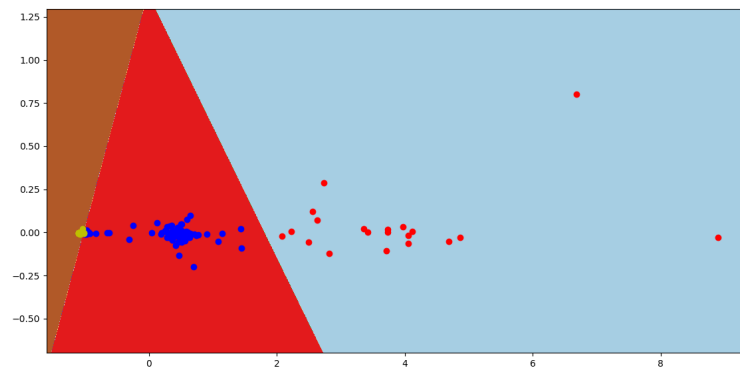


Figura 55: Resultado do modelo regressão logística múltipla.

6.3.2.1 Resultados dos testes

Para que fosse possível avaliar o modelo, foram recolhidos os mesmos sinais utilizados no teste anterior, sendo possível concluir que o modelo apresentou uma precisão de 94%. Sendo assim, pode-se concluir que os resultados obtidos foram semelhantes aos resultados do modelo "K-Mean", mas este modelo falhou mais vezes. A seguir, vão ser apresentadas as situações de falha com maior frequência.

A Figura 56 e 57 apresentam casos de falha com maior frequência, uma vez que o modelo considerou o componente do sinal inválidos como sendo um sinal válido.

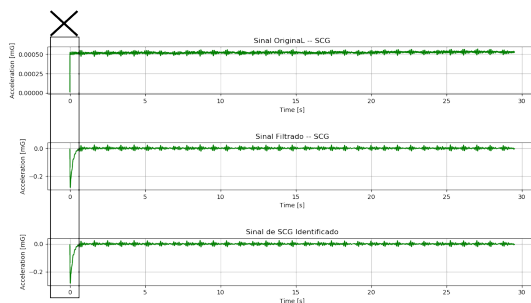


Figura 56: Teste falhado

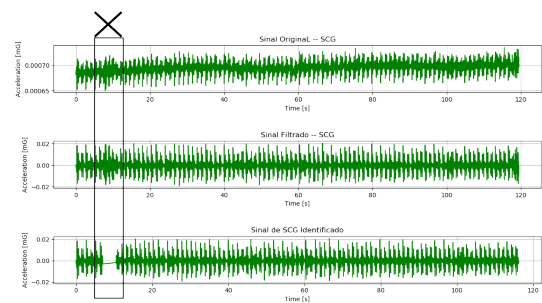


Figura 57: Teste falhado

6.3.3 Discussão dos resultados

Algoritmo	Accuracy	Precisão
Regressão logística múltipla	98%	94%
K-Mean	99%	97%

Tabela 1: Regressão logística múltipla vs K-Mean

Analisando os resultados obtidos na tabela 1, é possível concluir que o modelo "k-Mean" apresenta bons resultados, sendo capaz de identificar quase todos sinais válidos e com uma precisão aceitável. O modelo regressão logística múltipla não foi capaz de alcançar resultados melhores que o modelo "K-Mean" e a sua precisão também é menor que a precisão do modelo "K-Mean". Pode-se assim concluir que o modelo "k-Mean" é o mais adequado para a identificação de sinais da SCG. Além disso, apesar de apresentar algumas falhas, o modelo "K-Mean" conseguiu remover o sinal inválido em todos os testes.

6.4 Análise da SCG

Nesta secção apresentam-se os resultados da determinação dos parâmetros a partir dos sinais proveniente do dispositivo. Desde modo, foi realizado um método que extrai as características do sinal da SCG e os algoritmos de análise para determinação de métricas.

6.4.1 Determinação dos picos A0

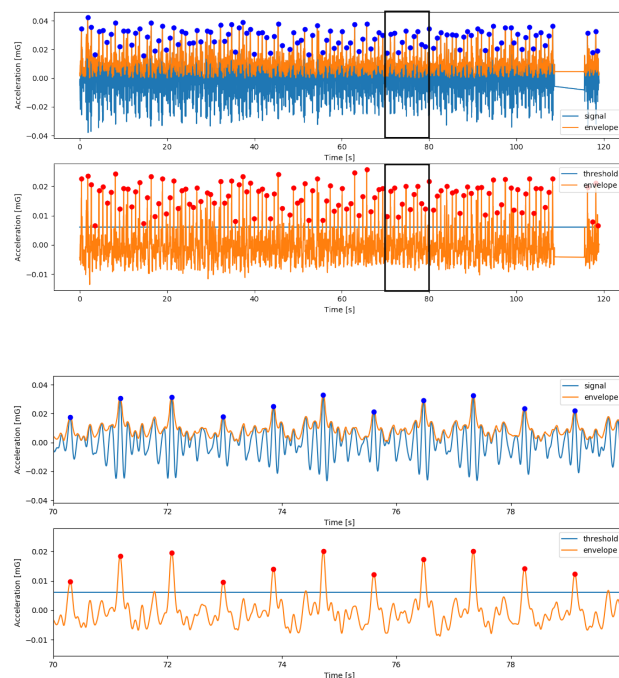


Figura 58: Análise do sinal da SCG adquirido e determinação dos picos A0

A Figura 58 apresenta o resultado da determinação dos picos A0. Para que fosse possível determinar os picos de maior amplitude (pico A0) para cada pulso, foi aplicado o método envelope para obter modulação do sinal. A partir desta modulação, foi definido um valor de "threshold" baseado no valor médio do módulo do sinal e nos máximos. Com base nisso, foi determinado o ponto de maior amplitude para cada pulso onde

o sinal envolvente é superior ao valor de "threshold" calculado.

De acordo com os picos AO determinados e o tempo da medição, é possível determinar a frequência cardíaca do paciente.

6.4.2 Determinação do sinal da respiração

De forma a extrair o sinal respiratório a partir do sinal apresentado na Figura 58, foi aplicado ao sinal um filtro passa banda para o intervalo de 0.05Hz – 1Hz, visto o componente de baixa frequência corresponder ao ciclo respiratório. O sinal resultante da aplicação desse filtro está representado na Figura 59.

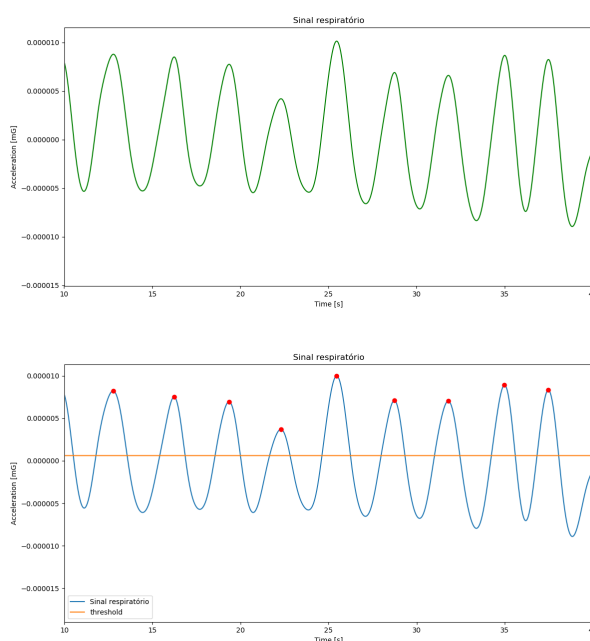


Figura 59: Sinal correspondente ao ciclo da respiração e determinação da frequência respiratória

Através da determinação dos picos de maior amplitude entre cada ciclo do sinal respiratório e do tempo de medição, torna-se possível calcular a frequência respiratória do paciente.

6.4.3 Ondas do sinal da SCG (Ensemble dos Sinais)

Para que fosse possível realizar uma análise mais concreta ao conjunto de ondas que caracterizam o sinal da SCG permitindo a identificação dos parâmetros, assim como a amplitudes e intervalos de tempo de propagação, foi realizada a sobreposição dos sinais e calculada a média do sinal, a qual pode ser observada na Figura 60.

A sobreposição dos vários intervalos identificados pelo pico AO e o cálculo do valor médio do conjunto de cada intervalo sobreposto determinam um sinal que corresponde à média desse sinal da SCG.

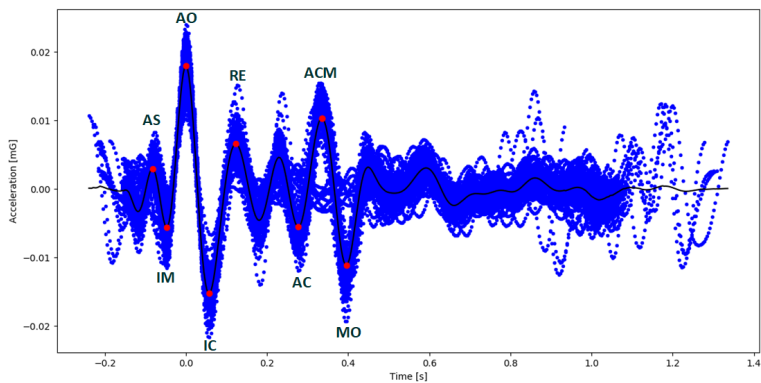


Figura 60: Ensemble dos sinais da SCG adquiridos.

Pela observação da Figura 60, é possível identificar a onda AO de maior amplitude e as restantes ondas secundárias (AS, IM, RE, AC, ACM, MO).

	A	B			
1	AS_IM[s]	0.072052728	23	RE_AC[s]	0.304586532
2	AS_AO[s]	0.140830332	24	RE_ACM[s]	0.307861656
3	AS_IC[s]	0.21288306	25	RE_MO[s]	0.31113678
4	AS_RE[s]	0.281660664	26	AC_ACM[s]	0.003275124
5	AS_AC[s]	0.586247196	27	AC_MO[s]	0.006550248
6	AS_ACM[s]	0.58952232	28	ACM_MO[s]	0.003275124
7	AS_MO[s]	0.592797444	29	AS[g]	0.009793547
8	IM_AO[s]	0.068777604	30	IM[g]	-0.02003104
9	IM_IC[s]	0.140830332	31	AO[g]	0.028369153
10	IM_RE[s]	0.209607936	32	IC[g]	-0.02302864
11	IM_AC[s]	0.514194468	33	RE[g]	0.0112978
12	IM_ACM[s]	0.517469592	34	AC[g]	-0.00545493
13	IM_MO[s]	0.520744716	35	ACM[g]	-0.00542171
14	AO_IC[s]	0.072052728	36	MO[g]	-0.00546448
15	AO_RE[s]	0.140830332	37	Heartrate - HR	64.74
16	AO_AC[s]	0.445416864	38		
17	AO_ACM[s]	0.448691988	39		
18	AO_MO[s]	0.451967112	40		
19	IC_RE[s]	0.068777604	41		
20	IC_AC[s]	0.373364136	42		
21	IC_ACM[s]	0.37663926	43		
22	IC_MO[s]	0.379914384	44		

Figura 61: Parâmetros calculados

A Figura 61 apresenta os parâmetros calculados a partir do sinal apresentado na Figura 60. Pela análise, pode concluir que a partir dos picos AO, ainda é possível obter bastante informação, como já foi referido anterior, o intervalo dos tempos entre as ondas típicas da sinal SCG pode ter algum valor significado e pode também ser utilizada para avaliar eventos do coração.

6.4.4 Análise da correlação entre ECO e ECG

Como descrito no capítulo 2, é possível existir uma relação entre os dados do ECO e da SCG e para que fosse possível realizar um estudo sobre essa correlação, foi aplicado o algoritmo "Pearson" que avaliava um possível grau da correlação linear entre esses dois tipos de dados.

No estudo, através dos resultados obtidos foi possível determinar alguma correlação entre os dados da SCG e do ECO, mas a maioria dos resultados obtidos apresenta uma correlação desprezável. Em seguida, pretende-se mostrar alguns resultados representativos que podem ser observados nas Figuras 62, 63, 64 e 65.

Pela observação das figuras, é possível observar que o gráfico apresentado em cada figura mostra 5 retas. A reta verde apresenta os dados dos pacientes AS, a reta roxa apresenta os dados dos pacientes "Healthy", a reta vermelha apresenta os dados dos pacientes HCM, a reta azul apresenta os dados dos pacientes DCM, a reta amarela apresenta os dados dos pacientes MI e a reta preta apresenta o grau da correlação entre estas retas. Além disso, o eixo y do gráfico apresenta os dados do ECO e o eixo x do gráfico apresenta os dados da SCG.

A Figura 62 apresenta um gráfico que mostra uma relação desprezável entre o "End-diastolic diameter" e o intervalo do tempo IM_MO da SCG, pois a intensidade da relação entre eles é 0.03 que está dentro da gama da correlação desprezável. Além disso, pela análise da reta preta apresentada no gráfico, é possível ver que o grau da correlação linear é basicamente zero, pois a reta preta é quase uma linha plana.

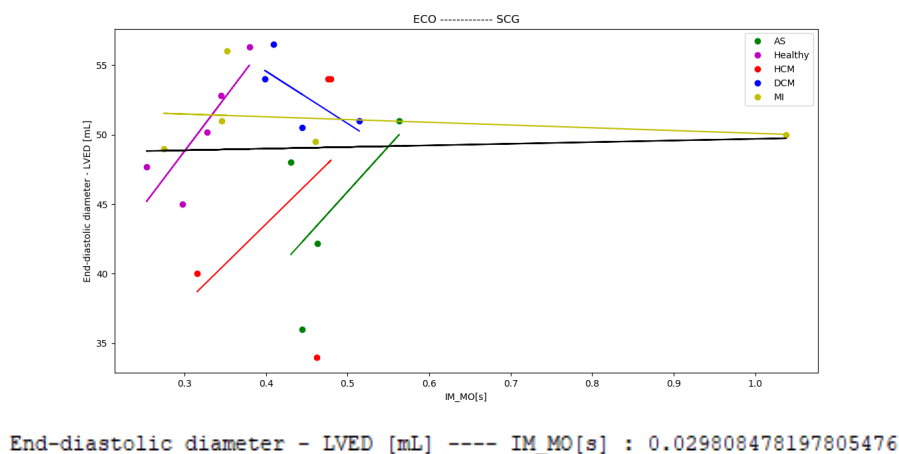


Figura 62: A correlação entre End-diastolic diameter e IM_MO

O gráfico apresentado na Figura 63 mostra uma relação fraca entre a "Ejection Fraction" e a amplitude IM da SCG, pois a intensidade da relação entre eles é 0.35 que está dentro da gama da correlação fraca. E pela análise da reta preta neste gráfico, é possível observar que o grau da correlação linear entre eles também é fraco, pois a reta preta não é uma Linha de bissecção.

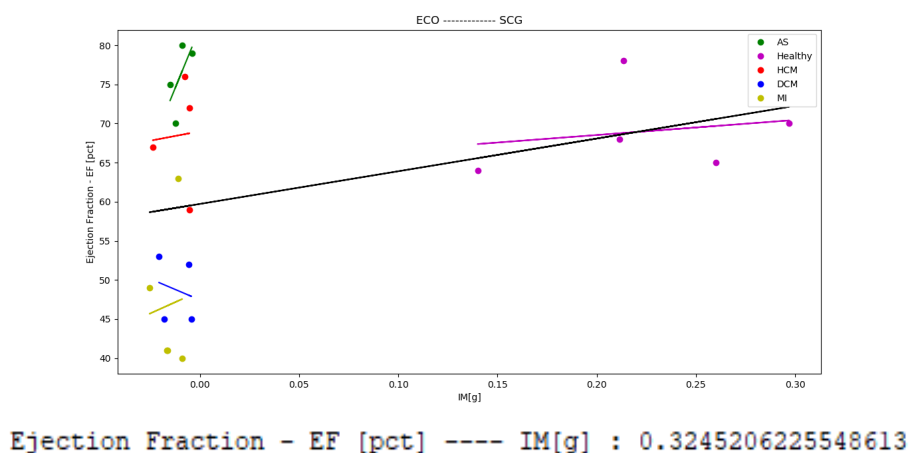


Figura 63: A correlação entre Ejection Fraction e IM

A Figura 64 apresenta um gráfico sobre o grau da relação entre o volume de "Stroke" e a amplitude AO da SCG. Pela análise do mesmo, é possível observar que existe uma correlação moderada entre eles, pois a intensidade da relação entre eles é 0.543 que está dentro da gama da correlação moderada. Aliás a reta preta obtida no gráfico aproxima a linha da bissecção.

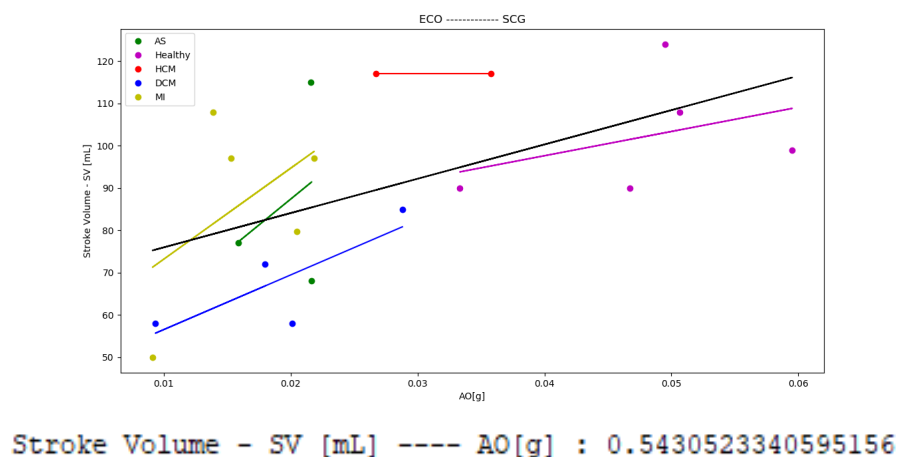


Figura 64: A correlação entre Stroke Volume e AO

A Figura 65 mostra uma relação forte entre o tempo da ejeção do sangue e o intervalo do tempo AO_ACM da SCG, pois a intensidade da relação entre eles é 0.744 que está dentro da gama da correlação forte. Pela observação do gráfico apresentado, é possível ver que o grau da correlação linear entre eles também é forte, pois a reta preta é quase uma linha da bissecção.

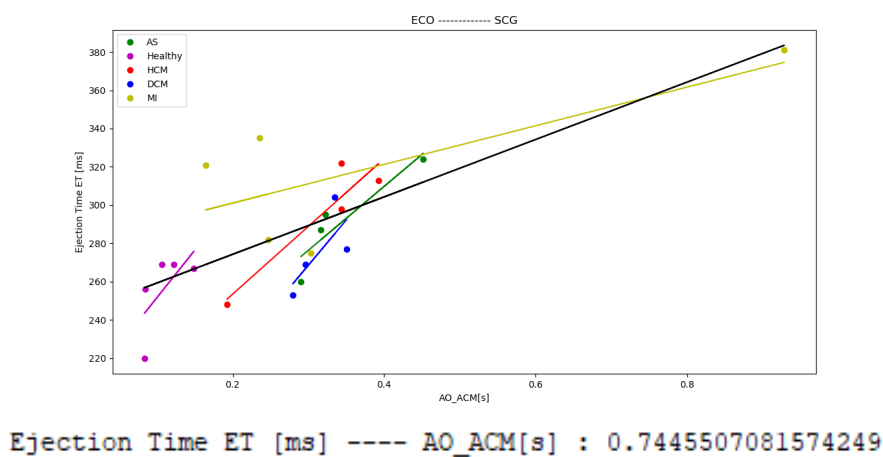


Figura 65: A correlação entre Ejection Time e AO_ACM

De acordo com os resultados obtidos, é possível concluir que não existe muita correlação entre os dados da SCG e do ECO com base deste estudo realizado, pois a maioria dos resultados obtidos apresenta uma relação desprezável e poucos apresentam uma relação fraca, moderada e forte.

6.5 Resultado da identificação da patologia cardíaca

De forma a identificar patologias cardíacas, foram utilizados os algoritmos SVM e DNN para criar os modelos de "Machine Learning". Os algoritmos SVM e DNN foram simulados no "Personal Computer", usando a ferramenta oferecida pela biblioteca "Tensorflow". Para que fosse possível avaliar e criar os modelos, foram realizados com a base de dados criada em ambiente hospitalar. E com o intuito de averiguar a qualidades dos dados oferecidos pelo sinal da SCG e do ECO, estes dois tipos de dados foram utilizados como entrada dos modelos de "Machine Learning" para investigar qual deles é mais adequado para aprendizagem de "Machine Learning" e identificação de doenças cardíaca.

6.5.1 Resultados com dados da SCG

Diversos sinais da SCG foram recolhidos em ambiente hospitalar, onde o paciente se sentou ao lado do dispositivo da SCG e os sinais medidos foram armazenados numa base de dados. Devido a algumas limitações, apenas foram extraídos os sinais provenientes de cinco tipos diferentes de pacientes: HCM, DCM, MI, AS e "Healthy".

O modelo foi criado primeiro usando o algoritmo DNN e depois o algoritmo SMV, no qual foram obtidos os melhores resultados, como pode ser visto na tabela 2. A tabela 2 representa os valores de "accuracy" dos modelos SVM e DNN baseados na base de dados criada. Analisando os resultados obtidos para cada um dos modelos, é possível ver que o resultado obtido pelo modelo SVM é perfeito, pois a "accuracy" do modelo é de 100%. Em relação ao modelo DNN, o modelo teve o pior desempenho, uma vez que a "accuracy" do modelo é de 71.43%.

Algoritmo	Accuracy
DNN	71.43%
SMV	100.00%

Tabela 2: DNN vs SMV baseados na base de dados da SCG

Para escolher o melhor valor para o parâmetro C do SVM, foram testados diferentes valores, e concluiu-se que o valor -0,5 será o melhor para o parâmetro. No algoritmo usado para o modelo DNN, foram construídas 3 camadas DNN com 10, 20, 10 unidades, respetivamente.

6.5.2 Resultados com dados do ECO

Os testes com os dados do ECO, foram realizados em ambiente hospitalar. Sendo os parâmetros recolhidos muito inferiores em quantidade comparativamente com os calculados a partir do sinal da SCG.

A tabela 3 apresenta os valores da "accuracy" dos modelos SVM e DNN baseados na base de dados do ECO criado. Pela análise do mesmo, é possível ver que a "accuracy" de ambos os modelos ficou acima 65%, mas o modelo SVM teve um desempenho melhor do que o método DNN.

Algoritmo	Accuracy
DNN	70.00%
SMV	100.00%

Tabela 3: DNN vs SMV baseados na base de dados do ECO

Para escolher o melhor valor para o parâmetro C do SVM tendo como base esta base de dados, foram testados diferentes valores e concluiu-se que -0,1 seria o melhor valor. Para o modelo DNN, foram construídas 3 camadas DNN com 20, 10, 20 unidades, respetivamente.

6.5.3 Discussão dos resultados

Analizando os resultados obtidos nas duas bases de dados, é possível concluir que o modelo SVM apresentou bons resultados. O modelo DNN nestas duas fases de dados não obteve resultados tão bons como o SVM. Analizando os resultados do desempenho do modelo DNN, o seu desempenho diminuiu na base de dados do ECO. Como o objetivo do trabalho era desenvolver um modelo capaz de identificar patologias cardíacas, analisando os resultados obtidos nestes testes, é possível concluir que isso é realizado com o modelo de SVM.

Devido a uma quantidade de dados insuficiente, só se conseguiu determinar os valores de "accuracy" dos modelos, de forma a determinar uma precisão dos modelos desenvolvidos, torna-se essencial a concretização de uma base de dados mais extensa. Assim, considera-se fundamental a aplicação dos modelos e a determinação das precisões baseadas num número mais alargado de pacientes, de forma a investigar a capacidade de diagnóstico dos modelos. Tal é um tópico de trabalho futuro para esta dissertação.

7 Conclusão e trabalho futuro

A implementação de um algoritmo para recolher e filtrar sinais proveniente do dispositivo da SCG, assim como a determinação de métricas de análise a partir dos sinais recolhidos, a correlação entre ECO e SCG e a implementação de um modelo de “Machine Learning” para estimar os estados de coração cardíacas, foram objetivos cumpridos com sucesso no desenvolvimento desta dissertação.

O sistema é baseado num dispositivo da SCG para detetar o movimento do coração e recolher o sinal da SCG. Como o sinal contém algum componente indesejado, torna-se necessário aplicar algoritmos dos filtros e “Machine Learning” (“K-Mean” e Regressão logística múltipla) para obter sinal válido. Em seguida, é necessário extrair as seguintes características: ondas AS, IM, AO, IC, RE, AC, ACM e MO, e aplicar um algoritmo para calcular os parâmetros necessários a partir dessas características. Usando os parâmetros calculados, identifica a correlação com dados do ECO. Depois utilizando esses parâmetros e os dados do ECO, os modelos de “Machine Learning” (SVM e DNN) foram treinados para reconhecer esses padrões e classificar o paciente como HCM, DCM, AS, MI e “Healthy”.

É importante referir que o caso de investigação realizada permitiu verificar a possibilidade de eliminação dos componentes inválidos ou ruídos contidos no sinal proveniente do dispositivo da SCG através do método Transformada Discreta “Wavelet” e o algoritmo “K-Mean”, e de determinação dos parâmetros baseando nestes sinais. Além disso, estima-se que pelos parâmetros obtidos e estando averiguada a relação dos valores destes com os dados do ECO, seja possível determinar alguma correlação como o tempo da ejeção do sangue com o intervalo do tempo AO_ACM da SCG e o volume de “Stroke” com a amplitude AO da SCG.

Para avaliar os modelos (SVM e DNN) desenvolvidos, dois tipos de treino foram realizados com dois bancos de dados diferentes, um com dados da SCG, outro com dados do ECO. Os resultados com esses dois bancos de dados geralmente tiveram uma exatidão melhor usando o modelo SVM. Na base de dados apenas com SCG, o modelo SVM obteve uma exatidão de 100.00%, contra 71.43% para o DNN. Em relação ao banco de dados com apenas ECO, o modelo SVM apresentou uma exatidão de 100.00% superando o modelo DNN que teve uma exatidão de 70.00%. Logo é permitido concluir que o modelo SVM é mais adequado para identificar as patologias cardíacas.

Com base nos resultados obtidos, foi possível concluir que para uma avaliação mais consistente da precisão de diagnóstico de patologias cardíacas dos modelos de “Machine Learning” torna-se necessária a realização de uma análise a um conjunto mais alargado de pacientes e testes. Para que seja realizada uma análise mais precisa sobre a precisão de diagnósticos dos modelos desenvolvidos, torna-se essencial a comparação de resultados a partir de diagnósticos de médico.

Em relação ao trabalho futuro, pretende-se realizar uma análise para perceber se o sistema desenvolvido

no decorrer desta dissertação pode ser usado para diagnóstico e antes disto, é necessário também extrair mais dados, tanto SCG como ECO, para determinar a precisão dos modelos desenvolvidos, porque pela identificação desta necessidade do sistema, são necessários mais dados para a realização de uma prova de conceito da análise de patologias cardíacas baseado em “Machine Learning”.

Referências

- [1] "OMS: 17,5 milhões de pessoas morrem todos os anos de doenças cardiovasculares | Agência Brasil - Últimas notícias do Brasil e do mundo." [Online]. Available: <http://agenciabrasil.ebc.com.br/geral/noticia/2016-09/oms-175-milhoes-de-pessoas-morrem-todos-os-anos-de-doencas-cardiovasculares>
- [2] "Fatores de Risco Para Doenças Cardíacas | ABC da Saúde." [Online]. Available: <https://www.abcdasaude.com.br/cardiologia/fatores-de-risco-para-doencas-cardiacas>
- [3] "Enfarte." [Online]. Available: <https://advancecare.pt/glossario/enfarte>
- [4] "OPAS/OMS Brasil - Doenças cardiovasculares." [Online]. Available: https://www.paho.org/bra/index.php?option=com_content&view=article&id=5253:doencas-cardiovasculares&Itemid=839
- [5] "Saúde do coração: como evitar as doenças que mais matam no mundo | GaúchaZH." [Online]. Available: <https://gauchazh.clicrbs.com.br/saude/vida/noticia/2017/09/saude-do-coracao-como-evitar-as-doencas-que-mais-matam-no-mundo-9901011.html>
- [6] "Doenças cardiovasculares matam 35 mil portugueses por ano – Observador." [Online]. Available: <https://observador.pt/2017/09/29/doencas-cardiovasculares-matam-35-mil-portugueses-por-ano/>
- [7] "Do que as pessoas morrem no mundo? - Unidades imaginárias." [Online]. Available: <http://unidadesimaginarias.com.br/index.php/2018/03/01/do-que-as-pessoas-morrem-no-mundo/>
- [8] "Por ano morrem 4 mil pessoas de enfarte - Sociedade - Correio da Manhã." [Online]. Available: <https://www.cmjornal.pt/sociedade/detalhe/por-ano-morrem-4-mil-pessoas-de-enfarte>
- [9] "Mais de 30 mil idosos em Portugal sofrem de estenose aórtica - News Farma." [Online]. Available: <http://www.newsfarma.pt/noticias/5571-mais-de-30-mil-idosos-em-portugal-sofrem-de-estenose-aortica.html>
- [10] "Cardiomiopatia dilatada - Distúrbios do coração e dos vasos sanguíneos - Manual MSD Versão Saúde para a Família." [Online]. Available: <https://www.msdmanuals.com/pt/casa/disturbios-do-coracao-e-dos-vasos-sanguineos/cardiomiopatia/cardiomiopatia-dilatada>
- [11] "Cardiomiopatia hipertrófica - Distúrbios do coração e dos vasos sanguíneos - Manual MSD Versão Saúde para a Família." [Online]. Available: <https://www.msdmanuals.com/pt/casa/disturbios-do-coracao-e-dos-vasos-sanguineos/cardiomiopatia/cardiomiopatia-hipertrofica>

- [12] K. Tavakolian, "CHARACTERIZATION AND ANALYSIS OF SEISMOCARDIOGRAM FOR ESTIMATION OF HEMODYNAMIC PARAMETERS," Tech. Rep., 2010. [Online]. Available: <https://core.ac.uk/download/pdf/56376301.pdf>
- [13] A. P. Colossimo, F. De Assis Costa, A. R. P. Riera, M. T. N. Bombig, V. C. Lima, F. A. H. Fonseca, M. C. O. Izar, B. L. Filho, D. Souza, and R. M. S. Pova, "Sensibilidade do Eletrocardiograma na Hipertrofia Ventricular de Acordo com Gênero e Massa Cardíaca Electrocardiogram Sensitivity in Left Ventricular Hypertrophy According to Gender and Cardiac Mass," Tech. Rep. [Online]. Available: <http://www.arquivosonline.com.br>
- [14] R. Póvoa and D. De Souza, "ARTIGO DE REVISÃO 81 Análise crítica do eletrocardiograma e do ecocardiograma na detecção da hipertrofia ventricular esquerda Critical review of the electrocardiogram and echocardiogram on the diagnosis of the left ventricular hypertrophy," Tech. Rep. 2, 2008. [Online]. Available: <http://departamentos.cardiol.br/dha/revista/15-2/09-eletrocardiograma.pdf>
- [15] "Teste ergométrico: o que é, para que serve e como é feito o exame?" [Online]. Available: <https://telemedicinamorsch.com.br/blog/teste-ergometrico-ou-teste-de-esforco>
- [16] B. J. Mortazavi, N. S. Downing, E. M. Bucholz, K. Dharmarajan, A. Manhapra, S. X. Li, S. N. Negahban, and H. M. Krumholz, "Analysis of Machine Learning Techniques for Heart Failure Readmissions," *Circulation: Cardiovascular Quality and Outcomes*, vol. 9, no. 6, pp. 629–640, nov 2016.
- [17] N. Gupta, N. Ahuja, S. Malhotra, A. Bala, and G. Kaur, "Intelligent heart disease prediction in cloud environment through ensembling," *Expert Systems*, vol. 34, no. 3, p. e12207, jun 2017. [Online]. Available: <http://doi.wiley.com/10.1111/exsy.12207>
- [18] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," *Proceedings of the Ninth Australasian Data Mining Conference - Volume 121*, no. December, pp. 23–30, 2011.
- [19] A. K. Pandey, P. Pandey, K. L. Jaiswal, and A. K. Sen, "A Heart Disease Prediction Model using Decision Tree," Tech. Rep. 6. [Online]. Available: www.iosrjournals.org
- [20] N. T. Alshohybe and Y. H. Ghaleb Al-Gaphari, "Diagnosis for Heart Disease Using Neural Network," Tech. Rep. 6. [Online]. Available: <https://ijcsits.org/papers/vol6no62016/3vol6no6.pdf>
- [21] I. Aec, M. Akhil Jabbar, B. L. Deekshatulu, & Priti, C. M. Akhil Jabbar, and P. Chandra, "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection," Type: Double Blind

- Peer Reviewed International Research Journal Publisher: Global Journals Inc, vol. 13, 2013. [Online]. Available: <https://pdfs.semanticscholar.org/072c/b3f9e9cbef46ed576edad28fed74d351954e.pdf>
- [22] S. Alty, "Cardiovascular disease prediction using support vector machines," *Circuits and Systems* ..., no. August 2015, pp. 376–379, 2003. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs/_all.jsp?arnumber=1562297
- [23] C. Technology, "Classification Of Heart Disease Using Svm And ANN," vol. 2, no. 9, pp. 694–701, 2013.
- [24] M. M. K. Dharani and C. Poovitha, "A DATA MINING MODEL TO PREDICT THE RISK OF HEART DISEASE USING MULTINOMIAL LOGISTIC REGRESSION (MLR)," *Tech. Rep.*, 2017. [Online]. Available: <http://www.ijser.org>
- [25] "Introduction to Random Forests." [Online]. Available: <https://www.datascience.com/resources/notebooks/random-forest-intro>
- [26] O. Pualy, "Random Forests for Medical Applications," 2012. [Online]. Available: <https://mediatum.ub.tum.de/doc/1094727/document.pdf>
- [27] J. A. Bulusu Deekshatulu, "Intelligent heart disease prediction system using random forest and evolutionary approach Intelligent heart disease prediction system using random forest and evolutionary approach," no. January, 2016.
- [28] E. Miranda, E. Irwansyah, A. Y. Amelga, M. M. Maribondang, and M. Salim, "Detection of Cardiovascular Disease Risk's Level for Adults Using Naive Bayes Classifier." *Healthcare informatics research*, vol. 22, no. 3, pp. 196–205, jul 2016. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/27525161><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4981580>
- [29] I. K. A. Enriko, M. Suryanegara, and D. Gunawan, "Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient ' s Health Parameters," vol. 8, no. 12, 1843.
- [30] A. Trevino, "Introduction to K-means Clustering," pp. 1–7, 2016. [Online]. Available: <https://www.datascience.com/blog/k-means-clustering>
- [31] "Algoritmo de Análise de Agrupamento k-means." [Online]. Available: <http://fp2.com.br/blog/index.php/2013/algoritmo-de-k-means/>
- [32] A. K. Pandey, P. Pandey, K. L. Jaiswal, and &. K. L. Jaiswal, "Classification Model for the Heart Disease Diagnosis," *Type: Double Blind Peer Reviewed International Research Journal Publisher:*

- Global Journals Inc, vol. 14, 2014. [Online]. Available: <https://pdfs.semanticscholar.org/c043/0451c075551c5e30c55bb050e09bceeed0c9.pdf>
- [33] Z. Cömert and A. F. Kocamaz, "Comparison of Machine Learning Techniques for Fetal Heart Rate Classification," vol. 132, 2017. [Online]. Available: <http://przyrbwn.icm.edu.pl/APP/PDF/132/app132z3p012.pdf>
- [34] "Using GPUs | TensorFlow." [Online]. Available: https://www.tensorflow.org/tutorials/using_gpu
- [35] "TensorFlow Essencial." [Online]. Available: <https://matheusfacure.github.io/2017/05/12/tensorflow-essencial/>
- [36] "Introdução ao Apache Mahout." [Online]. Available: <https://www.ibm.com/developerworks/br/java/library/j-mahout/index.html>
- [37] "Mahout Introduction." [Online]. Available: https://www.tutorialspoint.com/mahout/mahout_introduction.htm
- [38] "Open Source Machine Learning Libraries: TensorFlow, Theano, Torch, scikit-learn, Caffe." [Online]. Available: <https://www.altexsoft.com/blog/datascience/choosing-an-open-source-machine-learning-framework-tensorflow-theano-torch-scikit-learn-caffe/>
- [39] "Introduction - Accord.NET Machine Learning in C#." [Online]. Available: <http://accord-framework.net/intro.html>
- [40] P. K. Jain and A. K. Tiwari, "An algorithm for automatic segmentation of heart sound signal acquired using seismocardiography," 2016 International Conference on Systems in Medicine and Biology, ICSMB 2016, no. January, pp. 157–161, 2017.
- [41] A. Akhbardeh, K. Tavakolian, V. Gurev, T. Lee, W. New, B. Kaminska, and N. Trayanova, "Comparative analysis of three different modalities for characterization of the seismocardiogram." Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, vol. 2009, pp. 2899–903, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19964786>
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3313588>
- [42] A. Laurin, F. Khorsow, A. Blaber, K. Tavakolian, A. Laurin, F. Khorsow, and A. Blaber, "Accurate and consistent automatic seismocardiogram annotation without concurrent ECG Accurate and consistent automatic seismocardiogram annotation without concurrent ECG Accurate and

- consistent automatic seismocardiogram annotation without concurrent ECG2.” [Online]. Available: <https://hal.inria.fr/hal-01356419>
- [43] G. Shafiq, S. Tatinati, W. T. Ang, and K. C. Veluvolu, “Automatic Identification of Systolic Time Intervals in Seismocardiogram OPEN,” Nature Publishing Group, 2016. [Online]. Available: www.nature.com/scientificreports
- [44] D. Britto, F. Filho, J. Alexandre, and S. J. III, “Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r),” Tech. Rep. 1, 2009. [Online]. Available: <https://periodicos.ufpe.br/revistas/politica hoje/article/viewFile/3852/3156>
- [45] M. M. Mukaka, “Statistics corner: A guide to appropriate use of correlation coefficient in medical research.” Malawi medical journal : the journal of Medical Association of Malawi, vol. 24, no. 3, pp. 69–71, sep 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23638278><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3576830>

A ANEXOS

A.1 Correlação entre ECO e SCG

Stroke Volume - SV [mL]	-----	AS_IM[s]	-0.29526769269956094		
Stroke Volume - SV [mL]	-----	AS_AO[s]	-0.2376403752351146		
Stroke Volume - SV [mL]	-----	AS_IC[s]	-0.18048892078159207		
Stroke Volume - SV [mL]	-----	AS_RE[s]	-0.08473743364190349		
Stroke Volume - SV [mL]	-----	AS_AC[s]	0.24189983342780094		
Stroke Volume - SV [mL]	-----	AS_ACM[s]	0.2559632803518497		
Stroke Volume - SV [mL]	-----	AS_MO[s]	0.18894895474397855		
Stroke Volume - SV [mL]	-----	IM_AO[s]	0.3832891374894703		
Stroke Volume - SV [mL]	-----	IM_IC[s]	-0.2582415953597464		
Stroke Volume - SV [mL]	-----	IM_RE[s]	-0.17731076901390602		
Stroke Volume - SV [mL]	-----	IM_AC[s]	0.12116611412127276		
Stroke Volume - SV [mL]	-----	IM_ACM[s]	0.22662814694490244		
Stroke Volume - SV [mL]	-----	IM_MO[s]	0.21697097823662645		
Stroke Volume - SV [mL]	-----	AO_IC[s]	0.39769654020289574		
Stroke Volume - SV [mL]	-----	AO_RE[s]	0.43748385983259414		
Stroke Volume - SV [mL]	-----	AO_AC[s]	0.07300717599175142		
Stroke Volume - SV [mL]	-----	AO_ACM[s]	0.0901786428389654		
Stroke Volume - SV [mL]	-----	AO_MO[s]	0.1684553856377646		
Stroke Volume - SV [mL]	-----	IC_RE[s]	0.41703871879492754		
Stroke Volume - SV [mL]	-----	IC_AC[s]	0.3835605662822883		
Stroke Volume - SV [mL]	-----	IC_ACM[s]	0.4239201613506697		
Stroke Volume - SV [mL]	-----	IC_MO[s]	0.03742419103243658		
Stroke Volume - SV [mL]	-----	RE_AC[s]	0.322451535952971		
Stroke Volume - SV [mL]	-----	RE_ACM[s]	0.33818706072188737		
Stroke Volume - SV [mL]	-----	RE_MO[s]	0.27781305696597297		
Stroke Volume - SV [mL]	-----	AC_ACM[s]	0.4123586120658219		
Stroke Volume - SV [mL]	-----	AC_MO[s]	0.22262272050968843		
Stroke Volume - SV [mL]	-----	ACM_MO[s]	0.39794607464036813		
Stroke Volume - SV [mL]	-----	AS[g]	0.3814090769148011		
Stroke Volume - SV [mL]	-----	IM[g]	0.39000772743091533		
Stroke Volume - SV [mL]	-----	AO[g]	0.5430523340595156		
Stroke Volume - SV [mL]	-----	IC[g]	0.2578912462496557		
Stroke Volume - SV [mL]	-----	RE[g]	0.42146137568975245		
Stroke Volume - SV [mL]	-----	AC[g]	-0.20006940659118874		
Stroke Volume - SV [mL]	-----	ACM[g]	0.40057459636046355		
Stroke Volume - SV [mL]	-----	MO[g]	0.3965027019492695		

Figura 66: Correlação entre SV e ondas de SCG

Pre-ejection Period - PEP [ms]	-----	AS_IM[s]	0.1466135834471378		
Pre-ejection Period - PEP [ms]	-----	AS_AO[s]	0.1060130563952266		
Pre-ejection Period - PEP [ms]	-----	AS_IC[s]	0.0908927636005109		
Pre-ejection Period - PEP [ms]	-----	AS_RE[s]	0.10666779601142962		
Pre-ejection Period - PEP [ms]	-----	AS_AC[s]	-0.15398580945078982		
Pre-ejection Period - PEP [ms]	-----	AS_ACM[s]	-0.18173434158984064		
Pre-ejection Period - PEP [ms]	-----	AS_MO[s]	-0.1102491710377025		
Pre-ejection Period - PEP [ms]	-----	IM_AO[s]	-0.40945316721838204		
Pre-ejection Period - PEP [ms]	-----	IM_IC[s]	0.30210676802805997		
Pre-ejection Period - PEP [ms]	-----	IM_RE[s]	0.2686784009126159		
Pre-ejection Period - PEP [ms]	-----	IM_AC[s]	-0.03528931272734521		
Pre-ejection Period - PEP [ms]	-----	IM_ACM[s]	-0.13579803088870926		
Pre-ejection Period - PEP [ms]	-----	IM_MO[s]	-0.12567263719661806		
Pre-ejection Period - PEP [ms]	-----	AO_IC[s]	-0.41160925183556885		
Pre-ejection Period - PEP [ms]	-----	AO_RE[s]	-0.40545388119089665		
Pre-ejection Period - PEP [ms]	-----	AO_AC[s]	0.01876742109204958		
Pre-ejection Period - PEP [ms]	-----	AO_ACM[s]	-0.009624527365775728		
Pre-ejection Period - PEP [ms]	-----	AO_MO[s]	-0.06940424272197662		
Pre-ejection Period - PEP [ms]	-----	IC_RE[s]	-0.40788817833604685		
Pre-ejection Period - PEP [ms]	-----	IC_AC[s]	-0.295467210750213		
Pre-ejection Period - PEP [ms]	-----	IC_ACM[s]	-0.3355539824295194		
Pre-ejection Period - PEP [ms]	-----	IC_MO[s]	0.041140437628162156		
Pre-ejection Period - PEP [ms]	-----	RE_AC[s]	-0.24263708078271767		
Pre-ejection Period - PEP [ms]	-----	RE_ACM[s]	-0.2723431990254307		
Pre-ejection Period - PEP [ms]	-----	RE_MO[s]	-0.2044706008522844		
Pre-ejection Period - PEP [ms]	-----	AC_ACM[s]	-0.42427889797511187		
Pre-ejection Period - PEP [ms]	-----	AC_MO[s]	-0.16023097307628856		
Pre-ejection Period - PEP [ms]	-----	ACM_MO[s]	-0.3722486700449297		
Pre-ejection Period - PEP [ms]	-----	AS[g]	-0.37492892406796396		
Pre-ejection Period - PEP [ms]	-----	IM[g]	-0.39238461013616227		
Pre-ejection Period - PEP [ms]	-----	AO[g]	-0.4433902898443832		
Pre-ejection Period - PEP [ms]	-----	IC[g]	-0.3260782602093459		
Pre-ejection Period - PEP [ms]	-----	RE[g]	-0.4256978093681953		
Pre-ejection Period - PEP [ms]	-----	AC[g]	0.08177694130562249		
Pre-ejection Period - PEP [ms]	-----	ACM[g]	-0.3495390423705928		
Pre-ejection Period - PEP [ms]	-----	MO[g]	-0.38163847306749965		

Figura 67: Correlação entre PEP e ondas de SCG

Ejection Fraction - EF [pct]	-----	AS_IM[s]	-0.24983288836441428				
Ejection Fraction - EF [pct]	-----	AS_AO[s]	-0.16512217440206656				
Ejection Fraction - EF [pct]	-----	AS_IC[s]	-0.151985742225133				
Ejection Fraction - EF [pct]	-----	AS_RE[s]	-0.15089732437911899				
Ejection Fraction - EF [pct]	-----	AS_AC[s]	0.08549441643202012				
Ejection Fraction - EF [pct]	-----	AS_ACM[s]	0.10075038451482606				
Ejection Fraction - EF [pct]	-----	AS_MO[s]	0.04432049501689378				
Ejection Fraction - EF [pct]	-----	IM_AO[s]	0.3131259378637568				
Ejection Fraction - EF [pct]	-----	IM_IC[s]	-0.20856952429393963				
Ejection Fraction - EF [pct]	-----	IM_RE[s]	-0.208938448649875				
Ejection Fraction - EF [pct]	-----	IM_AC[s]	0.005273923005316872				
Ejection Fraction - EF [pct]	-----	IM_ACM[s]	0.0804194726846859				
Ejection Fraction - EF [pct]	-----	IM_MO[s]	0.0691054325958119				
Ejection Fraction - EF [pct]	-----	AO_IC[s]	0.317174103226059				
Ejection Fraction - EF [pct]	-----	AO_RE[s]	0.31437109635908167				
Ejection Fraction - EF [pct]	-----	AO_AC[s]	-0.0346348412806801				
Ejection Fraction - EF [pct]	-----	AO_ACM[s]	-0.019202625451409563				
Ejection Fraction - EF [pct]	-----	AO_MO[s]	0.02631465414717767				
Ejection Fraction - EF [pct]	-----	IC_RE[s]	0.3094843957156242				
Ejection Fraction - EF [pct]	-----	IC_AC[s]	0.20265100931254101				
Ejection Fraction - EF [pct]	-----	IC_ACM[s]	0.2289665581920002	Ejection Fraction - EF [pct]	-----	AS[g]	0.2803854651436721
Ejection Fraction - EF [pct]	-----	IC_MO[s]	-0.05754536073194579	Ejection Fraction - EF [pct]	-----	IM[g]	0.3245206225548613
Ejection Fraction - EF [pct]	-----	RE_AC[s]	0.16828167504889757	Ejection Fraction - EF [pct]	-----	AO[g]	0.37520452612612054
Ejection Fraction - EF [pct]	-----	RE_ACM[s]	0.18488263054461165	Ejection Fraction - EF [pct]	-----	IC[g]	0.2101422400895677
Ejection Fraction - EF [pct]	-----	RE_MO[s]	0.12898860707041304	Ejection Fraction - EF [pct]	-----	RE[g]	0.33390378635616585
Ejection Fraction - EF [pct]	-----	AC_ACM[s]	0.32467934424691913	Ejection Fraction - EF [pct]	-----	AC[g]	0.09144948487788901
Ejection Fraction - EF [pct]	-----	AC_MO[s]	0.1035248474026586	Ejection Fraction - EF [pct]	-----	ACM[g]	0.26359934689320336
Ejection Fraction - EF [pct]	-----	ACM_MO[s]	0.2898380393371701	Ejection Fraction - EF [pct]	-----	MO[g]	0.28548374374949975

Figura 68: Correlação entre EF e ondas de SCG

Interventricular Septal thickness -IVSd [mm]	-----	AS_IM[s]	0.18337904116341847				
Interventricular Septal thickness -IVSd [mm]	-----	AS_AO[s]	0.43396859501557616				
Interventricular Septal thickness -IVSd [mm]	-----	AS_IC[s]	0.4535264890246555				
Interventricular Septal thickness -IVSd [mm]	-----	AS_RE[s]	0.3691614741569528				
Interventricular Septal thickness -IVSd [mm]	-----	AS_AC[s]	-0.12151638765558029				
Interventricular Septal thickness -IVSd [mm]	-----	AS_ACM[s]	-0.14507901164068895				
Interventricular Septal thickness -IVSd [mm]	-----	AS_MO[s]	-0.16014260292440297				
Interventricular Septal thickness -IVSd [mm]	-----	IM_AO[s]	0.5858027389883025				
Interventricular Septal thickness -IVSd [mm]	-----	IM_IC[s]	0.5596101026537368				
Interventricular Septal thickness -IVSd [mm]	-----	IM_RE[s]	0.38413542628434233				
Interventricular Septal thickness -IVSd [mm]	-----	IM_AC[s]	-0.13797737547397476				
Interventricular Septal thickness -IVSd [mm]	-----	IM_ACM[s]	-0.1621072480723729				
Interventricular Septal thickness -IVSd [mm]	-----	IM_MO[s]	-0.17733035996491975				
Interventricular Septal thickness -IVSd [mm]	-----	AO_IC[s]	0.25357327994323997				
Interventricular Septal thickness -IVSd [mm]	-----	AO_RE[s]	0.05758975570918616				
Interventricular Septal thickness -IVSd [mm]	-----	AO_AC[s]	-0.17629991787757845				
Interventricular Septal thickness -IVSd [mm]	-----	AO_ACM[s]	-0.2002504452581749				
Interventricular Septal thickness -IVSd [mm]	-----	AO_MO[s]	-0.21518354989142022				
Interventricular Septal thickness -IVSd [mm]	-----	IC_RE[s]	-0.165488915704444				
Interventricular Septal thickness -IVSd [mm]	-----	IC_AC[s]	-0.18609080376748255				
Interventricular Septal thickness -IVSd [mm]	-----	IC_ACM[s]	-0.2102203972565518	Interventricular Septal thickness -IVSd [mm]	-----	AS[g]	-0.3020899198066576
Interventricular Septal thickness -IVSd [mm]	-----	IC_MO[s]	-0.22544723670692154	Interventricular Septal thickness -IVSd [mm]	-----	IM[g]	0.20726840523213738
Interventricular Septal thickness -IVSd [mm]	-----	RE_AC[s]	-0.1781347774356048	Interventricular Septal thickness -IVSd [mm]	-----	AO[g]	0.18770657234326266
Interventricular Septal thickness -IVSd [mm]	-----	RE_ACM[s]	-0.20244660814167112	Interventricular Septal thickness -IVSd [mm]	-----	IC[g]	-0.497759453647269
Interventricular Septal thickness -IVSd [mm]	-----	RE_MO[s]	-0.2182683022271779	Interventricular Septal thickness -IVSd [mm]	-----	RE[g]	0.46083374048191766
Interventricular Septal thickness -IVSd [mm]	-----	AC_ACM[s]	-0.18726169118788907	Interventricular Septal thickness -IVSd [mm]	-----	AC[g]	-0.0027811329344286555
Interventricular Septal thickness -IVSd [mm]	-----	AC_MO[s]	-0.22290398932850414	Interventricular Septal thickness -IVSd [mm]	-----	ACM[g]	-0.16036383758070683
Interventricular Septal thickness -IVSd [mm]	-----	ACM_MO[s]	-0.18521793292247915	Interventricular Septal thickness -IVSd [mm]	-----	MO[g]	-0.05132331104652841

Figura 69: Correlação entre IVSd e ondas de SCG

PP [mm]	-----	AS_IM[s]	-0.17143853486535487		
PP [mm]	-----	AS_AO[s]	-0.015752088931369958		
PP [mm]	-----	AS_IC[s]	-0.03130272716897194		
PP [mm]	-----	AS_RE[s]	-0.12428915285155835		
PP [mm]	-----	AS_AC[s]	-0.2599617349755346		
PP [mm]	-----	AS_ACM[s]	-0.2867574998386468		
PP [mm]	-----	AS_MO[s]	-0.3166597265773266		
PP [mm]	-----	IM_AO[s]	0.1906465410496986		
PP [mm]	-----	IM_IC[s]	0.1107338120022804		
PP [mm]	-----	IM_RE[s]	-0.046757818023607775		
PP [mm]	-----	IM_AC[s]	-0.24493184457936765		
PP [mm]	-----	IM_ACM[s]	-0.27184890906357706		
PP [mm]	-----	IM_MO[s]	-0.30216548628205503		
PP [mm]	-----	AO_IC[s]	-0.055533962819879025		
PP [mm]	-----	AO_RE[s]	-0.22106144947653886		
PP [mm]	-----	AO_AC[s]	-0.2548530434794977		
PP [mm]	-----	AO_ACM[s]	-0.28071746491115684		
PP [mm]	-----	AO_MO[s]	-0.3105729831413016		
PP [mm]	-----	IC_RE[s]	-0.35185277788018643		
PP [mm]	-----	IC_AC[s]	-0.24885785747149472		
PP [mm]	-----	IC_ACM[s]	-0.274712344746497	PP [mm]	-----
PP [mm]	-----	IC_MO[s]	-0.30488717675118904	PP [mm]	-----
PP [mm]	-----	RE_AC[s]	-0.2324445951200588	PP [mm]	-----
PP [mm]	-----	RE_ACM[s]	-0.2584338429697501	PP [mm]	-----
PP [mm]	-----	RE_MO[s]	-0.28938160771623894	PP [mm]	-----
PP [mm]	-----	AC_ACM[s]	-0.19238637684414578	PP [mm]	-----
PP [mm]	-----	AC_MO[s]	-0.3192124001866583	PP [mm]	-----
PP [mm]	-----	ACM_MO[s]	-0.3676255183382047	PP [mm]	-----
				AS[g]	-0.5150880667474472
				IM[g]	0.43270950255903917
				AO[g]	-0.12194341954196364
				IC[g]	-0.24882725268979936
				RE[g]	0.11454565347770201
				AC[g]	0.23737083627252437
				ACM[g]	-0.13272468568618723
				MO[g]	-0.11682696358661111

Figura 70: Correlação entre PP e ondas de SCG

Left Ventricle Mass - LVMass [g]	-----	AS_IM[s]	-0.15425507548934664		
Left Ventricle Mass - LVMass [g]	-----	AS_AO[s]	0.07929086516809598		
Left Ventricle Mass - LVMass [g]	-----	AS_IC[s]	-0.020195334498119638		
Left Ventricle Mass - LVMass [g]	-----	AS_RE[s]	-0.1632682641180788		
Left Ventricle Mass - LVMass [g]	-----	AS_AC[s]	-0.16461497681271373		
Left Ventricle Mass - LVMass [g]	-----	AS_ACM[s]	-0.22888882398868857		
Left Ventricle Mass - LVMass [g]	-----	AS_MO[s]	-0.2566079216844145		
Left Ventricle Mass - LVMass [g]	-----	IM_AO[s]	0.348492418761581		
Left Ventricle Mass - LVMass [g]	-----	IM_IC[s]	0.11250026185699694		
Left Ventricle Mass - LVMass [g]	-----	IM_RE[s]	-0.1148271624639956		
Left Ventricle Mass - LVMass [g]	-----	IM_AC[s]	-0.15102491064358528		
Left Ventricle Mass - LVMass [g]	-----	IM_ACM[s]	-0.2154048298962927		
Left Ventricle Mass - LVMass [g]	-----	IM_MO[s]	-0.24347087787469152		
Left Ventricle Mass - LVMass [g]	-----	AO_IC[s]	-0.2754011434108259		
Left Ventricle Mass - LVMass [g]	-----	AO_RE[s]	-0.44507214017422936		
Left Ventricle Mass - LVMass [g]	-----	AO_AC[s]	-0.17294375508774237		
Left Ventricle Mass - LVMass [g]	-----	AO_ACM[s]	-0.23617118883686064		
Left Ventricle Mass - LVMass [g]	-----	AO_MO[s]	-0.26378178850517925		
Left Ventricle Mass - LVMass [g]	-----	IC_RE[s]	-0.5324541401923613		
Left Ventricle Mass - LVMass [g]	-----	IC_AC[s]	-0.15752802304433697	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	IC_ACM[s]	-0.22000146240904292	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	IC_MO[s]	-0.24785301695351333	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	RE_AC[s]	-0.13343839129786575	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	RE_ACM[s]	-0.19565006675980467	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	RE_MO[s]	-0.22408974784647542	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	AC_ACM[s]	-0.5389965679040616	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	AC_MO[s]	-0.5430776029689359	Left Ventricle Mass - LVMass [g]	-----
Left Ventricle Mass - LVMass [g]	-----	ACM_MO[s]	-0.3394581544596389	Left Ventricle Mass - LVMass [g]	-----
				AS[g]	0.0935414062999638
				IM[g]	-0.18704976106275628
				AO[g]	0.10874249603966302
				IC[g]	0.09998859544237378
				RE[g]	-0.2972274635668331
				AC[g]	0.2709283632693886
				ACM[g]	0.012684669185167693
				MO[g]	-0.20897691654912295

Figura 71: Correlação entre LVMass e ondas de SCG

Figura 72: Correlação entre LVM e ondas de SCG

Figura 73: Correlação entre LVEd e ondas de SCG

Figura 74: Correlação entre EF e ondas de SCG

Maximum Flow Rate - MFR [m/s]	-----	AS_IM[s]	-0.02831328758103528	
Maximum Flow Rate - MFR [m/s]	-----	AS_AO[s]	0.08281930365350103	
Maximum Flow Rate - MFR [m/s]	-----	AS_IC[s]	0.10373634790164636	
Maximum Flow Rate - MFR [m/s]	-----	AS_RE[s]	0.11711254884806767	
Maximum Flow Rate - MFR [m/s]	-----	AS_AC[s]	0.0637668556854498	
Maximum Flow Rate - MFR [m/s]	-----	AS_ACM[s]	0.1116989027082786	
Maximum Flow Rate - MFR [m/s]	-----	AS_MO[s]	0.1566990428932242	
Maximum Flow Rate - MFR [m/s]	-----	IM_AO[s]	-0.1877621177897617	
Maximum Flow Rate - MFR [m/s]	-----	IM_IC[s]	0.2338728661396323	
Maximum Flow Rate - MFR [m/s]	-----	IM_RE[s]	0.2257538514713373	
Maximum Flow Rate - MFR [m/s]	-----	IM_AC[s]	0.11984084786713126	
Maximum Flow Rate - MFR [m/s]	-----	IM_ACM[s]	0.13823131013481196	
Maximum Flow Rate - MFR [m/s]	-----	IM_MO[s]	0.1610896744511986	
Maximum Flow Rate - MFR [m/s]	-----	AO_IC[s]	-0.1966122151760292	
Maximum Flow Rate - MFR [m/s]	-----	AO_RE[s]	-0.18707026746587624	
Maximum Flow Rate - MFR [m/s]	-----	AO_AC[s]	0.12537072602148727	
Maximum Flow Rate - MFR [m/s]	-----	AO_ACM[s]	0.16420906387245318	
Maximum Flow Rate - MFR [m/s]	-----	AO_MO[s]	0.16422306195510247	
Maximum Flow Rate - MFR [m/s]	-----	IC_RE[s]	-0.1958136294145267	
Maximum Flow Rate - MFR [m/s]	-----	IC_AC[s]	-0.015817794896645455	
Maximum Flow Rate - MFR [m/s]	-----	IC_ACM[s]	0.024775157742015816	
Maximum Flow Rate - MFR [m/s]	-----	IC_MO[s]	0.18899670025703066	
Maximum Flow Rate - MFR [m/s]	-----	RE_AC[s]	0.005430044298870188	
Maximum Flow Rate - MFR [m/s]	-----	RE_ACM[s]	0.05378765180124019	
Maximum Flow Rate - MFR [m/s]	-----	RE_MO[s]	0.10787062699298007	
Maximum Flow Rate - MFR [m/s]	-----	AC_ACM[s]	-0.12233208017155918	
Maximum Flow Rate - MFR [m/s]	-----	AC_MO[s]	0.283347907004592	
Maximum Flow Rate - MFR [m/s]	-----	ACM_MO[s]	-0.12006549989989429	
Maximum Flow Rate - MFR [m/s]	-----	AS[g]	-0.21777497411459945	
Maximum Flow Rate - MFR [m/s]	-----	IM[g]	-0.18000548537101174	
Maximum Flow Rate - MFR [m/s]	-----	AO[g]	-0.14291173333836105	
Maximum Flow Rate - MFR [m/s]	-----	IC[g]	-0.2567542733464848	
Maximum Flow Rate - MFR [m/s]	-----	RE[g]	-0.15253878252116285	
Maximum Flow Rate - MFR [m/s]	-----	AC[g]	-0.16472620269092997	
Maximum Flow Rate - MFR [m/s]	-----	ACM[g]	-0.1613225394801964	
Maximum Flow Rate - MFR [m/s]	-----	MO[g]	-0.16740124067081943	

Figura 75: Correlação entre MFR e ondas de SCG